Structure of the 40S ribosomal subunit from *Plasmodium falciparum* By Homology and *De novo* modeling

Harrison Ndung'u Mwangi (BSc. Kenyatta University)

Reg No: I56/61628/2011

Center for Biotechnology and Bioinformatics

University of Nairobi

A thesis submitted to the Board of Postgraduate Studies, University of Nairobi, in partial fulfillment for the award of

Master of Science in Bioinformatics

Declaration and approval

I declare that the work presented here in is original and has not been presented for award of any degree or at any other university to the best of my knowledge

Name: Harrison Ndung'u Mwangi

Registration: I56/61628/2011

Signature: _____Date_____

Approval

This thesis has been submitted for examination with our approval as the University Supervisors:

Prof, Francis Mulaa, PhD

Department of biochemistry &

Centre for Biotechnology and Bioinformatics

University of Nairobi

Signature: _____Date_____

Fredrick Sijenyi, PhD

Senior Scientist,

DNA Software, Inc.

Ann Arbor, Michigan

U.S.A

Signature: _____Date_____

Prof, Peter Waiganjo Wagacha, PhD

School of computing and informatics &

Centre for Biotechnology and Bioinformatics

University of Nairobi

Signature: _____Date_____

Acknowledgements

I acknowledge God Almighty for the strength and grace He bestowed upon me while undertaking this research study. I am highly indebted to my research supervisors: Dr. Fredrick Sijenyi for offering me an excellent opportunity to work with him. Fred, you have been a wonderful supervisor and an admirable mentor. I thank you, Prof. Francis Mulaa for your invaluable guidance in my research work, especially for offering constructive critique during the whole process of my study in the varsity. Many thanks to Prof. Peter WaiganjoWagacha for playing a vital role in fanning the flames of my curiosity in computational sciences. I also acknowledge the constant support from Prof. James Ochanda through whom I got the opportunity of studying in one of the finest Bioinformatics academic and research programs in the country. To my colleagues Peterson, Maureiq and Eric, thanks for the support during my time at graduate school. I am sincerely grateful to my family for constant support, unconditional love and encouragement throughout my whole life and most importantly during my research work. I am immeasurably grateful to my Mother Naomi Njeri for instilling in me a great sense of discipline and hard work; and for being an everyday source of inspiration. My siblings Dr Jefferson Wanyoike, Tobias Wanyoike and Stephen Chege- Thank you for encouraging me to pursue my dreams. Finally, I thank my Dad and mum for funding my two year study to undertake my MSc studies at the University of Nairobi.

Dedication

I dedicate this thesis to my lovely family: To James Mwangi Wanyoike and Naomi Njeri for seeing to it that I get the best education ever; to my siblings for making me smile even when the going gets tough; Maureen Michire, Simon Kuria and Dr G. Gachara for being an epitome of hope and excellence.

Contents

Declaration and approval…… ii
Acknowledgements…… iii
Dedication……...iv
LIST OF TABLES……vii
LIST OF FIGURES……..viii
ABSTRACT……x
LIST OF ABBREVIATION AND SYMBOLS ……..xi
Chapter One ……1
1.0 Introduction……1
1.1 Protein synthesis in eukaryotic ribosome ……..1
1.1.1 Initiation ……2
1.1.2 Elongation……...2
1.1.3 Termination……3
1.2 Eukaryote Ribosome modeling ……6
1.3 Prokaryotic vs. eukaryotic protein synthesis ……7
1.4 Problem statement……..8
1.5 Overall objective……8
1.5.1 Specific objectives ……..8
1.6 Rationale ……9
Chapter Two ……10
2.0 Eukaryotic Ribosome ……..10
2.1 40S Ribosomal Subunit ……...14
2.2 18S rRNA Structure ……17
2.3 RNA as a drug Target ……21
2.4 40S Ribosomal proteins……...23
2.5 Protein homology modeling……29
2.6 RNA homology modeling……30
Chapter Three……31
3.0 Methodology……31
3.1.Template selection ……...31
3.1.1 Alignment ……32

3.1.2Model building……...33
3.1.3 Side chain modeling……..34
3.1.4 Energy minimization……34
3.1.5 Structure Validation ……..35
3.2 Proof of concept ……..35
3.3 Homology /comparative modeling process ……38
3.4 Preliminary results……...39
3.5 Conclusion of the Preliminary results ……42
Chapter Four ……..43
4.0 Results ……43
4.1 Protein models of *Plasmodium falciparum* ……43
4.1.1 Plasmodium falciparum 40S proteins ……43
4.1.2 Protein structure Validation……...47
4.2 Plasmodium falciparum 18S r RNA ……49
Chapter Five……...53
5.0 Discussions and conclusion ……54
Chapter Six…….64
6.0 References……64
APPENDICES ……..74
7.0 Appendices ……..74
7.1 List of computer tools used in this work ……74
7.2 Sequences used as template ……80
7.3 Sequences of the 40S Plasmodium falciparum subunit used in this work ……85

LIST OF TABLES

Table 1 *Saccharomyces cerevisiae* 40S ribosomal subunit showing some proteins modeled with their RMSD. 40

Table 2: Summary of *Plasmodium falciparum* 40S ribosomal proteins models RMSDs and protein classification showing no of residues their homologous counterparts of the template)... 47

Table 3: RMSD of the *Plasmodium falciparum* protein models using Procheck a validation tool. 48

Table 4: *Plasmodium falciparum* 18S r RNA Energy optimization Table. 52

LIST OF FIGURES

Figure 1 Eukaryotic translation mechanism. 5

Figure 2 An oversimplified diagram showing the structure of the 80S Eukaryotic ribosome.....11

Figure 3 *S. cerevisiae* 80S ribosome crystal structure at 4.15Å (A). View of the E site, and (B) View from the A site Proteins and rRNA in the 40S are colored dark and light blue, respectively,. 13

Figure 4 Eukaryotic 40S subunit tertiary and secondary structures. (A). 18S rRNA domains are colored according to domain (5′ domain, red; central domain, green; 3′ major domain, yellow; 3′ minor domain, blue; ESs, magenta), and the proteins as gray cartoons. 16

Figure 5 Watson–Crick G-C and A-U base pairs. 18

Figure 6 RNA secondary structure motifs showing representation of Watson crick base paring. 19

Figure 7 Protein structure hierarchy. 25

Figure 8 Schematic diagram showing general systematic Homology /comparative modeling process..... 38

Figure 9 A and B Three Dimensional structures of some of the modeled proteins of the *Saccharomyces cerevisiae* with their RMSD compared to the crystalline models. 39

Figure 10 Diagram showing the two dimensional structure of *Saccharomyces cerevisiae* with modeled domains 3'Major and 3'Minor with their RMSDs..... 41

Figure 11: 3Dstructure of all ribosomal proteins of the 40S subunit of *Plasmodium falciparum*. Backbone of all the proteins as shown colored to the distribution (structures shown)..... 45

Figure 12: The modeled 3'Minor domain of *Plasmodium falciparum* superimposed with the template *Tetrahymena thermophila* 3'Minor domain. 49

Figure 13: 3'major model of *Plasmodium falciparum* superimposed with template *T thermophila* 49

Figure 14: Modeled central domain of *Plasmodium falciparum* superimposed with the template *Tetrahymena thermophila.* 50

Figure 15: 5'major domain model superimposed with template 50

Figure 16: The 18S rRNA structure of *Plasmodium falciparum* superimposed with the template *T thermophila* 51

Figure 17: *Plasmodium falciparum* 18S r RNA Energy optimization differences from the positive before energy minimization to a more favorable negative structure. 52

Figure 18 Architectural tertiary structure of plasmodium falciparum 40S front and back view. Shown is the 18S rRNA colored differently depending with domains. 53

Figure 19 18S rRNA Secondary structure of *Tetrahymena thermophila* and *Plasmodium falciparum* showing the four non-homologous regions between the two structures located at the central and 3'major domain. 60

Figure 20 shows the combined structure of the 18S rRNA and anchored to it are the 34 proteins.. 63

ABSTRACT

Generation of the three dimensional structures of macromolecules using *in silico* structural modeling technologies such as homology and *de novo* modeling has improved dramatically and increased the speed in which tertiary structures of organisms of interest can be generated. This is especially the case if a homologous crystal structure is already available. High resolution structures can be rapidly created using only their sequence information as input and thus increasing the speed of scientific discoveries. In this study, a host of homology modeling and structure prediction tools such as RNA123 and SWISS –MODEL among others, were used to generate the 40S subunit from *Plasmodium falciparum.* This structure was modeled using the published crystal structure from *Tetrahymena Thermophila*, a homologous eukaryote X-ray structure. In the absence of any information from the solved *Plasmodium falciparum* 40S ribosomal crystal structure, the model accurately depicts a global topology, secondary and tertiary connections, and gives an overall RMSD value of 3.9 Å relative to the templates crystal structure. The model accuracy is even better than prior hypothesis, though deviations are modestly larger for areas that had no homology between the templates. These results lay ground work for using this approach for larger and more complex eukaryotic ribosomes, as well for still larger RNAs, RNA-protein complexes and entire ribosomal subunits. The model created will provide a scaffold onto which *in silico* ligands screening can be performed with the ultimate goal of developing new classes of anti-malarial compounds.

Keywords: Ribosome; 40S subunit; RNA; structure; comparative analysis; three-dimensional modelling; RMS

LIST OF ABBREVIATION AND SYMBOLS

A-Adenine

BLAST- Basic Local Alignment Search Tool

C –Cytosine

DNA –Deoxy-Ribonucleic acid

E. coli- Escherichia coli

FRET-Fluorescence Resonance Energy Transfer FRET

G - Guanine

GDT_TS –Global Distance Test Total Score

GTP- Guanosine Tri-phosphate

MATCH-CHECK- Tool for matching proteins

Da- Daltons

MRNA –Messenger Ribonucleic Acid

NMR-structures- Nuclear magnetic resonance spectroscopy

PDB - protein data bank

RMSD -Root-Mean-Square Deviation

RNA –Ribonucleic Acid

rRNA – ribosomal Ribonucleic Acid

tRNA- Transfer Ribonucleic Acid

T-Thymine

U-Uracil

UV-Ultra-Violet

ES interactions- Expansion Segments

CSP-Constraint Space Programming

eIF- Eukaryotic initiation factor

Chapter One

1.0 Introduction

Ribosomes are cellular organelles found in the cytoplasm and primarily responsible for protein synthesis in the cell. Ribosomes' were first observed as dense particles or granules under an electron microscope (Frank 2009). The Eukaryote ribosome is a large complex (about 2.6 MDa) molecular machine composed of rRNAs and proteins (Garrett 1999; White 2000; Ramakrishnan 2009). In the past few years, a combination of X-ray crystallography, NMR spectroscopy and Cryo-electron microscopy has provided new data on the structure of ribosome's (Frank 1999). The Eukaryotic ribosome (80S) comprises of two subunits, a large subunit (60S) and small subunit (40S). The ribosome plays a major role during translation of RNA to the various proteins they code for. The process of translation occurs when the message contained in mRNA is decoded and the respective amino acids synthesized into a growing polypeptide chain which eventually folds into a three dimensional functional structure. Protein synthesis is critical for cell viability, hence highlighting the importance of the ribosome in the cell (Sauer 2007; Sauer 2007).

1.1 Protein synthesis in eukaryotic ribosome

The process of protein synthesis is performed by the ribosome which acts together with a number of addition factors to translate the genetic information encoded in the mRNA. The process of translation can be divided into three phases; initiation, elongation, and termination (Preiss and M 2003; Albert L. Lehninger, David L. Nelson et al. 2005).

1.1.1 Initiation

Initiation process starts with the dissociation of the vacant ribosome into its constituent subunits catalyzed by initiation factors eIF3 and eIFA (Klinge, Voigts-Hoffmann et al. 2012). Several initiation factors as well as the Met-tRNA bind to the dissociated 40S subunit to form a pre initiation complex which is then delivered by the G-protein to the P-site (Huang, Yoon et al. 1997; Hinnebusch and Lorsch 2012).The 40S subunit forms a ternary complex after joining with the eIF2, Met-tRNA and GTP (Preiss and M 2003; Lori A. Passmore 2007) (Figure 1).

This complex binds to a small ribosomal subunit occupying the active sites in the ribosomes which is known as the protein site (P site). The 5' end of the mRNA that is to be translated binds the initiation complex after its recognition which in turn slides down to the initiation codon which is always an AUG sequence of amino acids that now show that the mRNA is attached to the ribosome (Albert L. Lehninger, David L. Nelson et al. 2005). The A (acceptor) site also known as decoding site on the ribosome (Frank Schluenzen, k Ante Tocilj et al. 2000), is occupied by tRNA that ensures correct base paring occurs on the codon on the mRNA with its anticodon.

1.1.2 Elongation

During the elongation phase, actual protein synthesis takes place leading to the growth of the polypeptide chain (Figure 1). Here, formation of the peptide bond between the P site bound peptide and the A-site bound amino acid occurs catalyzed by RNA contacts in the large subunit(Doudna and Rath 2002; Ramakrishnan 2002)

The 40S complex is then joined by the large subunit and a second tRNA brought in the ribosome by an elongation factor (Maria Selmer, Christine M. Dunham et al. 2006). Peptidyl transferase center which is part of the large ribosomal subunit that mediates separation of the first amino acid from the tRNA and formation of a peptide bond between the initial methionine and the amino acid formed enhancing elongation (Ramakrishnan 2002; Kapp and Lorsch 2004; Jha and Komar 2011).The second amino acid in this position is adjacent to methionine and the uncharged tRNA molecule now occupies the P site (Preiss and M 2003; Maria Selmer, Christine M. Dunham et al. 2006). A process known as translocation thereafter begins where the ribosome moves down the mRNA by one codon. This movement in turn shifts the growing peptide chain to position P and the result is an empty A site where a new charged tRNA can enter the pair simply by forming a hydrogen bond between the codon and the anticodon. For stable binding to occur, tRNA will hold into place long enough. Previously the uncharged tRNA occupied at the P site is expelled out of the ribosome and will be recharged and recycled by the cell (Barbara S. Schuwirth, Maria A. Borovinskaya et al. 2005). The cycle of guanosine triphosphate (GTP) hydrolysis has been shown to provide a general mechanism of alteration in numerous functional control (Li and Zhang 2004; Grigorenko, Shadrina et al. 2008). Hydrolysis of guanosine triphosphate (GTP) provides all the required energy needed in the process. Continuation of the process goes on the entire mRNA, until the first stop codon is encountered (N. Ban, P. Nissen et al. 2000).

1.1.3 Termination

Termination in eukaryotes has been shown to be a GTP dependent process, where a polypeptide release factor oversees the hydrolysis of the last peptidyl-tRNA with the

release of the emerging polypeptide chain in the ribosome in presence of a three nucleotides known as termination, stop or nonsense codon and GTP (Zhouravleva, Frolova et al. 1995; Dunkle and Cate 2010). The translocation of a stop codon into a ribosomal A-site by action of the elongation factor EF-G or eEF2 triggers a pept-tRNA hydrolysis chemical reaction that occurs at the peptidyltransferase center in the large ribosomal subunit of the ribosome (Kisselev, Ehrenberg et al. 2003; MOORE and STEITZ 2003).The release factors eRF1 and eRF3 recognize the stop codon and stimulate hydrolysis of the completed polypeptide from the last P-site tRNA a mechanism that is not clear until recently (Beringer 2008; Dunkle and Cate 2010). Termination factor releases the completed protein from the last tRNA and the ribosome dissociates into its component parts (Ramakrishnan 2009). After termination proper three dimensional conformation of the protein must be achieved to ensure their biological active form coupled by the posttranslational processing which may happen before or after folding. Enzymatic processing of the newly synthesized polypeptide that include removal of one or more amino acids, addition of acetyl, phosphoryl, methyl, carboxyl or other groups to certain amino acid residues, proteolytic cleavage and attachment of oligosaccharides or prosthetic groups may occur (Albert L. Lehninger, David L. Nelson et al. 2005).

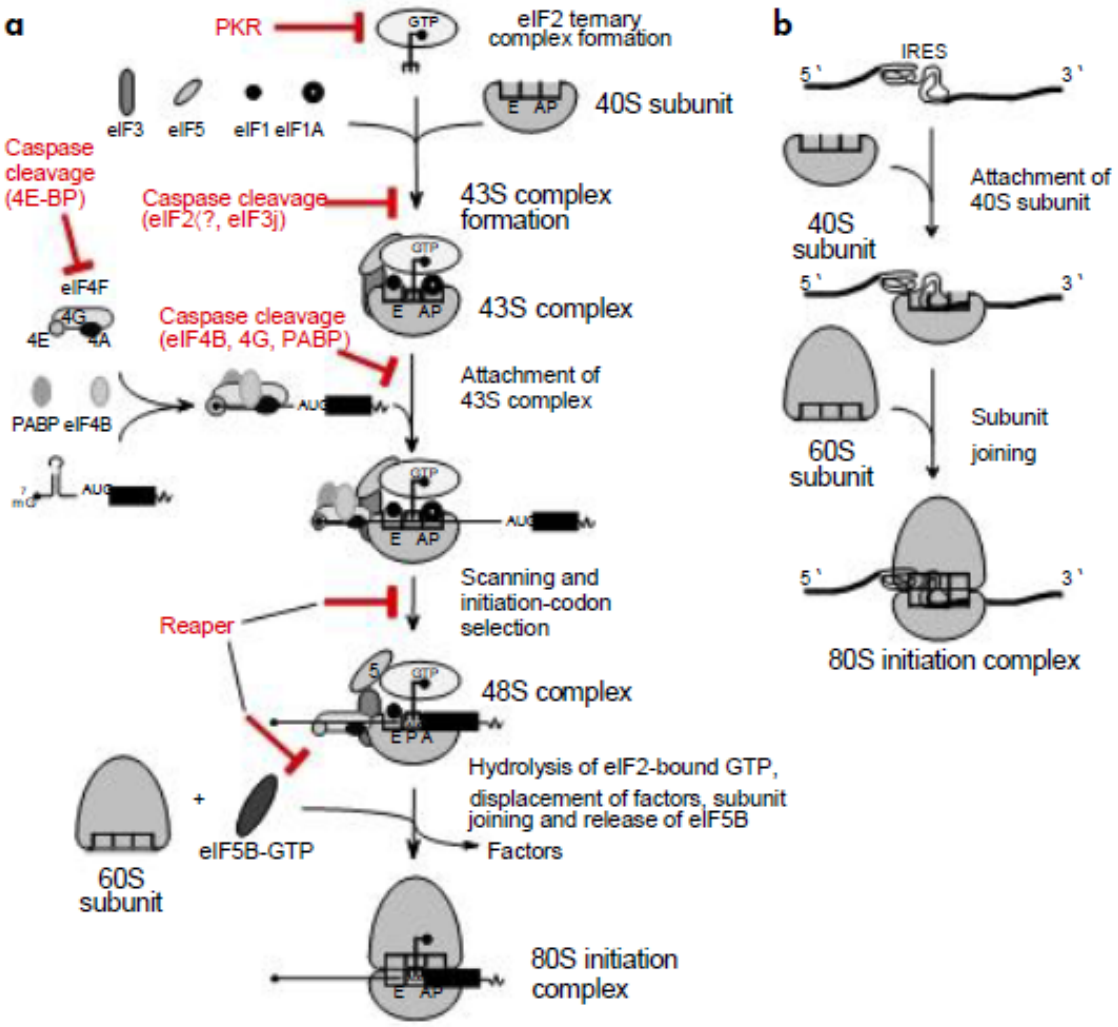


Figure 1 Eukaryotic translation mechanism (a). Detailed protein synthesis initiation process (b) Schematic diagram showing the process at which the two subunits come together to form the 80S during elongation (Hellen 2006).

1.2 Eukaryote Ribosome modeling

By using advanced structural modeling techniques such as electron microscopy and nuclear magnetic resonance, new information is being generated that has been helpful in understanding the complexity of the eukaryotic ribosome over the prokaryotic ribosome. One such group that has had success in trying to obtain an accurate structure of ribosome is led by Joachim Frank. As he puts it "With this working model we can now understand and predict more of the eukaryotic ribosome's biochemical interactions with other molecules in the cell" (Frank 2009). Ribosomal structural studies have been performed by different laboratories around the world for many years leading to the awarding of the 2009 Nobel Prize for Chemistry, to three leading researchers in the field, (HHMI investigator Thomas A. Steitz, Ada Yonath of the Weizmann Institute of Science in Israel, and Venkatraman Ramakrishnan of the MRC Laboratory of Molecular Biology in England). Generation of high resolution X-ray structures of eukaryotic 80S ribosomes lags behind compared to the bacteria ribosomes structures available due to their greater degree of complexity (Dinman 2009). The use of Cryo-electron microscopy and single-particle reconstruction have been used to determine structures of a transplanting plant 80S ribosome at 5.5Å resolution which map together with 6.1Å of *Saccharomyces cerevisiae* 80S ribosome. Recent work however, has led to the determination of the first X-ray structure of the yeast ribosome at 3.0 Å resolution. This has been fueled by the advancement in crystallographic methods for structural determination (Ben-Shem, Garreau de Loubresse et al. 2011). This revelation led to the present yeast 80S ribosomal structure determined by cryo-EM and X-ray crystallography procedures and further facilitated the comparison of ribosomes from different species and kingdoms of life

(Jenner, Melnikov et al. 2012). This has enabled the modeling of~98% of the rRNA and accurate assignments of its expansion segments and the variable regions reveling unique ES-ES and r-protein–ES interactions, providing insight into the structure and evolution of the eukaryotic ribosome (Jean-Paul Armachea, Alexander Jarascha et al. 2010). Recently developed technologies that would help obtain a detailed structure of eukaryote ribosomes using *in silico* modeling methods to generate the *P. falciparum* 40S subunit structure will be useful in understanding its biochemical interaction with other molecules in basic life forms.

1.3 Prokaryotic vs. eukaryotic protein synthesis

Eukaryotic ribosomes are considerably more complex and larger than their prokaryotic counterparts. Prokaryotic and eukaryotic proteins synthesis differs considerably at the level of initiation, though their core function is conserved. Eukaryotic protein synthesis process is initiated by the small ribosomal subunit (40S) binding initiation factors that facilitate scanning of the messenger RNA. The ribosome has been validated as an antibiotic target in the treatment of bacterial infections (Gilly, Benson et al. 1985; Qi, Wu et al. 2009; Andrew P. Carter, William M. Clemons et al. 2000). Thus high resolution structures that will be generated in this study for *Plasmodium falciparum* will offer new prospects for developing newer drugs that will combat the of increasing resistance of the parasite to current drugs. The 40S subunit from *Plasmodium falciparum* model provides insights into specific aspects of protein synthesis which would be very important to help understand more about the eukaryotic parasites that has been a major global challenge and find new intervention that will limit its pathogenesis.

1.4 Problem statement

Can we generate a three dimensional model of the 40s subunit from *P. falciparum* using a homologous crystal structure as a template and threading in the *P. falciparum* sequence? How reliable is this models in revealing active sties that can be used as targets for *in silico* structure based drug design?

1.5 Overall objective

Generate the structure of the 40S ribosomal subunit from *P. falciparum* through a combination of homology modeling and *de novo* structure prediction technologies.

1.5.1 Specific objectives

1. Determine high resolution homologous RNA and protein crystal structures that can be used as templates for the *P. falciparum* small ribosomal subunit.

2. Use the identified crystal structures as templates in homology modeling software and thread in the *P. falciparum* sequences as the query sequence

3. Merge the Protein and RNA models into a single complex and run energy minimization to determine the overall complex structure.

1.6 Rationale

Determining the first structure of the *P. falciparum* 40 S will lead to better understanding of the structural basis for its protein-synthesizing roles in the cell. This will enable researchers in the field of drug development to run *In silico* ligand screening experiments using the solved *P. falciparum* 40S structure as a target against a library of potential anti-malarial compounds. Drug leads identified through this method can lead to further biochemical and *In vitro* binding studies with the ultimate goal of developing new class of anti-malarial drugs. The use of structure prediction and modeling technologies in this study will dramatically reduce the time it takes from target identification to drug lead determination. Additionally, millions of compounds that can be tested *In silico* against the generated structure would have been impossible to test experimentally. This demonstrates the growing power of bioinformatics and three dimensional structural modeling software technologies.

Chapter Two

2.0 Literature Review

2.1 Eukaryotic Ribosome

A eukaryote ribosome is designated as 80S and contains two subunits. The smaller subunit 40s is comprised of 18s rRNA and 33 proteins whereas the large subunit 60S (Figure2) is comprised of the 28S, 5S and 5.8S rRNA and 49 proteins (Ben-Shem, Garreau de Loubresse et al. 2011; Jenner, Melnikov et al. 2012; Andrew P. Carter, William M. Clemons et al. 2000).

X-ray crystallography and Cryo-electron microscopy methods have been used to solve the three dimensional structures of the ribosome with or without complexed cofactors tRNA, mRNA among other macromolecules. successfully (Ditlev E. Brodersen 2000; Frank Schluenzen, k Ante Tocilj et al. 2000; N. Ban, P. Nissen et al. 2000; Ben-Shem, Garreau de Loubresse et al. 2011; Jenner, Melnikov et al. 2012).

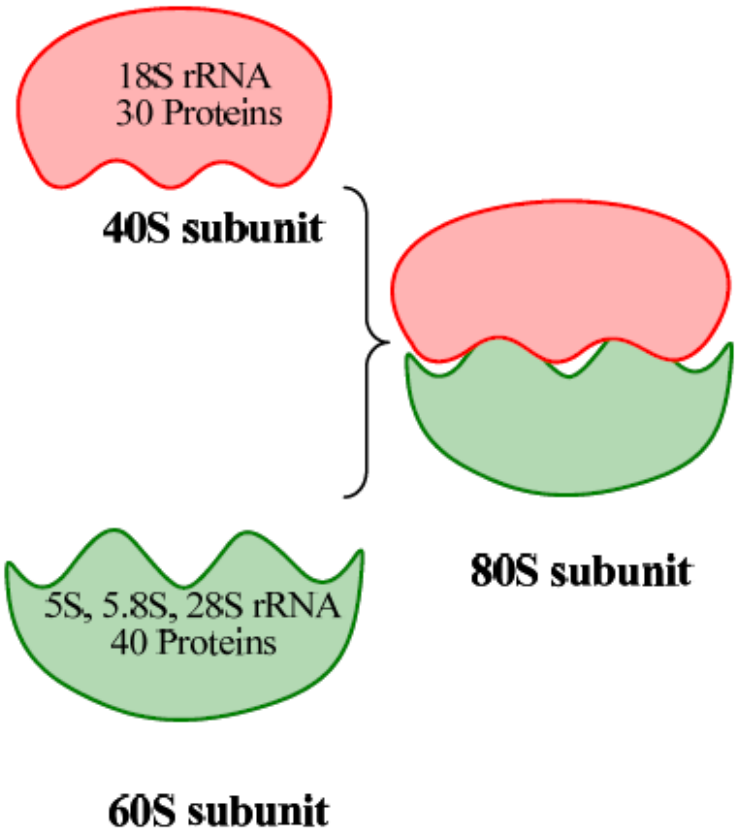


Figure 2 An oversimplified diagram showing the structure of the 80S Eukaryotic ribosome made up of the two subunits 40S and 60S and their composition that facilitate the process of translation.

First models of eukaryotic ribosome's at resolution between 6.1 and 15 Å were provided by Cryo- electron microscopy reveled the location and the shapes of the RNA expansion segments and indicated the position of additional protein moieties (Christian MT Spahn 2001; Christian MT Spahn 2004; Thomas Becker 2009). Later, a crystalline structure of the complete eukaryotic ribosome from *Saccharomyces cerevisiae* was determined at 4.15 Å and later 3.0 Å resolutions (Adam B en-Shem 2010; Jenner, Melnikov et al. 2012) shown in Figure 3. These structures with more clarity gave more insights on understanding the process of translation which captured the ratcheted states of the ribosome which had been postulated over 40 years (Bretscher 1968; Spirin 1969). Recent overall crystal structure of the eukaryotic ribosome of *S. cerevisiae* obtained (Figure 2.1) revealed basic architectural similarity but a larger assembly compared with the prokaryotic counterpart (Adam B en-Shem 2010). An addition to this structure is that it shows the E-site, A-site, the ribosomal proteins of both the 60S and the 40S subunits together with the expansion segments which gave more knowledge about eukaryotic protein synthesis process. This followed through earlier studies that were done and showed both interfaces of the 60S and 40S subunits views with numbered bridges (Yusupov Marat M, Gulnara Zh. Yusupova et al. 2001).

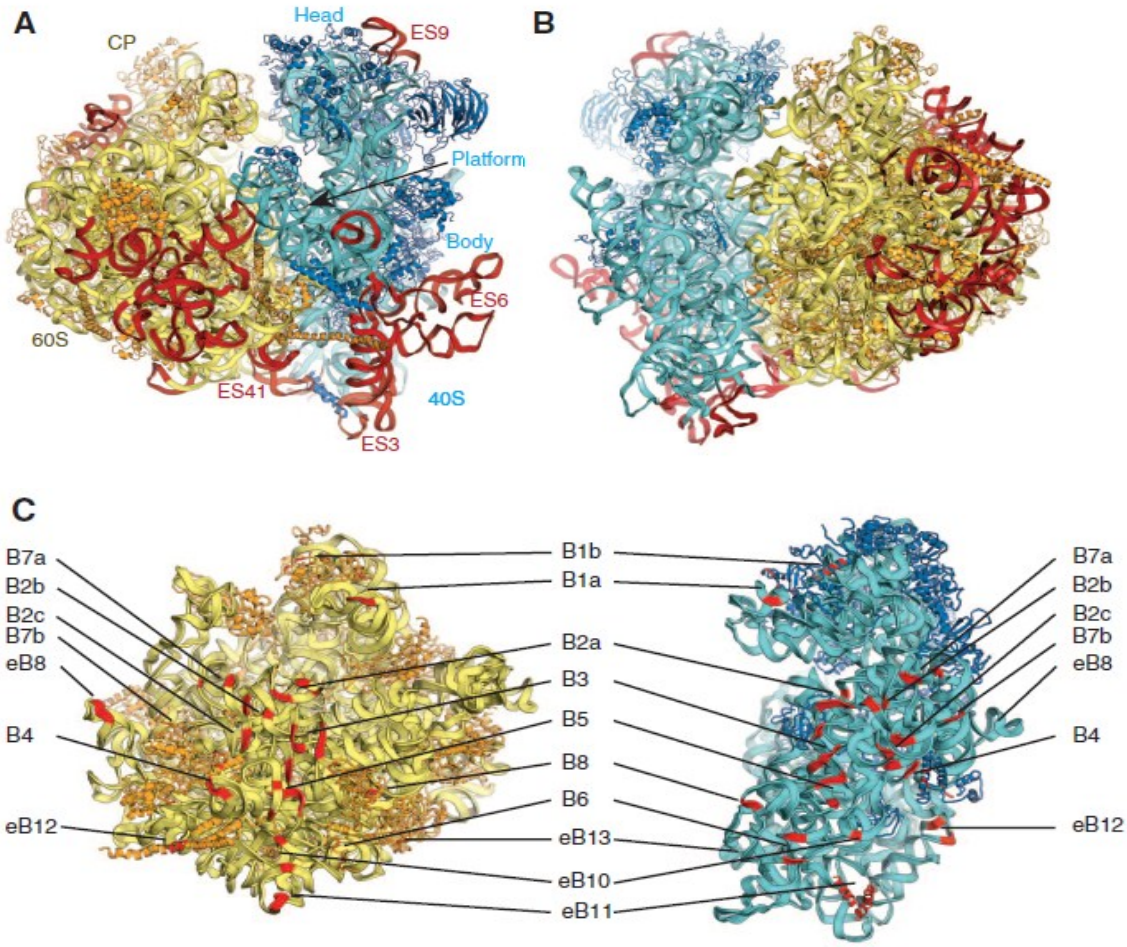


Figure 3 *S. cerevisiae* 80S ribosome crystal structure at 4.15Å (A). View of the E site, and (B) View from the A site Proteins and rRNA in the 40S are colored dark and light blue, respectively, and dark and pale yellow, respectively, in the 60S,expansion segments are in red (Adam B en-Shem 2010). (C) Views of the 60S and 40S subunits interfaces with bridges numbered essentially as in and colored red.

2.2 40S Ribosomal Subunit

The eukaryotic ribosome is highly complex with a molecular mass of ~4 million Da and roughly 40% larger as compared to 2.8 million Da of the prokaryotic ribosome (Bielka 1982; Adam B en-Shem 2010; Julius Rabl, Marc Leibundgut et al. 2011). The 60S subunit is larger consisting of 28S, 5S and 5.8S r RNA and ~40 proteins whereas the eukaryotic smaller subunit 40S is larger by almost 500 kDa compared to the prokaryotic small 30S ribosomal subunit (William M. Clemons Jr 1999; John Dresios 2006; Taylor, Devkota et al. 2009). The 60S subunit size varies ranging from 1.2 to 1.7 million Da due to the size variation of the 28S r RNA (Bielka 1982; Adriana Verschoor, Suman Srivastava et al. 1996). More recent publications have reported higher resolution structures of the eukaryotic 40S subunit at 3.0 Å (Julius Rabl, Marc Leibundgut et al. 2011). In the 40S subunit, the 18S rRNA has three notable sequence insertions, as compared to the 16S as shown in Figure 4 (Hendriks, Goris et al. 1991; Neefs, Van de Peer et al. 1991). These insertions are known as expansion segments (ES). ES3 is located in the 5'major domain, ES6 and ES7 in the central domain, ES9 in the 3'major and finally the ES12 located at the 3'minor domain illustrated in figure 3 (Julius Rabl, Marc Leibundgut et al. 2011). The eukaryote ribosome 40S subunit contains ~30 proteins, 18 of which do not have homologs in bacteria, and the 18S rRNA which comprises of ~45% mass of the subunit (Adriana Verschoor, Suman Srivastava et al. 1996; Julius Rabl, Marc Leibundgut et al. 2011). Recent publications of eukaryotic ribosome structures that have been obtained later show more defined interactions between the ribosomal RNA and the proteins. Such include the eukaryotic 40S ribosomal subunit of *Tetrahymena thermophila*

(Figure 4). This has provided high resolution structures revealing the atomic detail of individual functional state of the ribosome

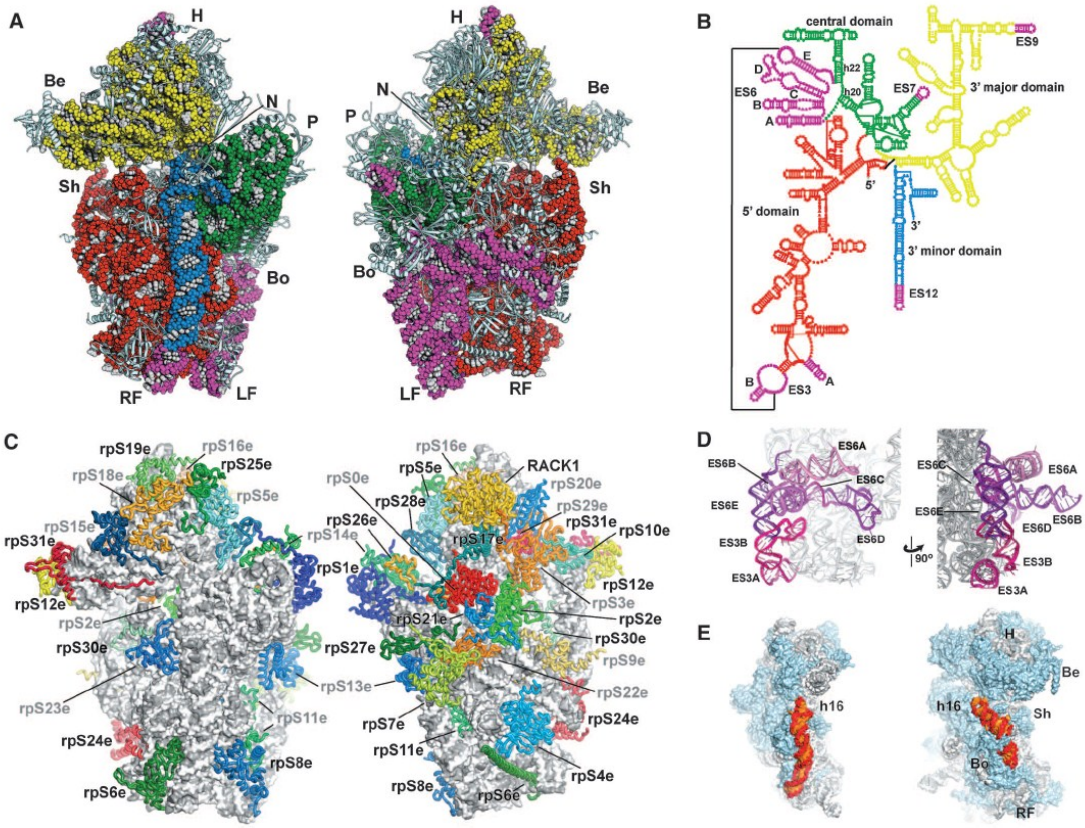


Figure 4 Eukaryotic 40S subunit tertiary and secondary structures. (A). 18S rRNA domains are colored according to domain (5′ domain, red; central domain, green; 3′ major domain, yellow; 3′ minor domain, blue; ESs, magenta), and the proteins as gray cartoons (abbreviations: H, head; Be, beak; N, neck; P, platform; Sh, shoulder; Bo, body; RF, right foot; LF, left foot).(B) secondary structure of 18S RNA of *T . thermophila* showing domains and ESs. (C) 40S proteins shown in cartoon in different colors; rRNA in gray. (D) View of ES6 and ES3 quaternary interactions. (Nygard 2004). (E) The position of helix h16 in bacterial 30S [left, adapted from (Andrew P. Carter, William M. Clemons et al. 2000)] and in 40S.

2.3 18S rRNA Structure

RNA molecules are polymers of nucleotides comprised of 3'-5' phosphor-diester linked ribose sugars attached to the four bases, pyrimidines: cytosine and uracil and purines: adenine and guanine in contrast with the DNA (Barbara S. Schuwirth, Maria A. Borovinskaya et al. 2005). RNA is single stranded containing ribose sugars and uracil base in place of thymine. Base pairing occurs in both DNA and RNA using the hydrogen bonds formed this pairing patterns are governed by Watson and crick rules (Crick 1953), where adenine complements uracil in RNA (thymine in DNA) and guanine with cytosine in both known as conical base pairing. In places where this pairing does not happen it's known as non-Watson crick structures among these are the sheared GA, GA imino, AU reverse Hoogsteen, and the GU and AC wobble pairs (Crick 1953; Fredrick Sijenyi, Pirro Saro et al. 2012; Westhof 2012) (Figure 5). Determining the RNA secondary structure is the first step of understanding its mechanism of action which is defined by the canonical base paring (Mathews and Turner 2006). The secondary structure of the RNA can adopt elements such as internal loops, mismatches bulges, multi-branched junctions, hairpins and pseudo knots (Figure 6). Presence of these diverse motifs that can be adopted by the RNA makes it play its functional roles such as recognition, interaction, metal binding and other enzymatic activities (Fredrick Sijenyi, Pirro Saro et al. 2012). The 18S fold primarily defines the eukaryotic 40S subunit structure which can be divided into features of the small ribosomal subunit including the head, platform, body, beak, shoulder, right foot and left foot (Julius Rabl, Marc Leibundgut et al. 2011) (Figure 4 (A)).The secondary structure of the eukaryotic ribosome forms a structure with four major domains

which are named according to the region of the sequence from 5’ to the 3’ end. They are 5’major, central, 3’major, and finally 3’minor (Figure 4 (B)).

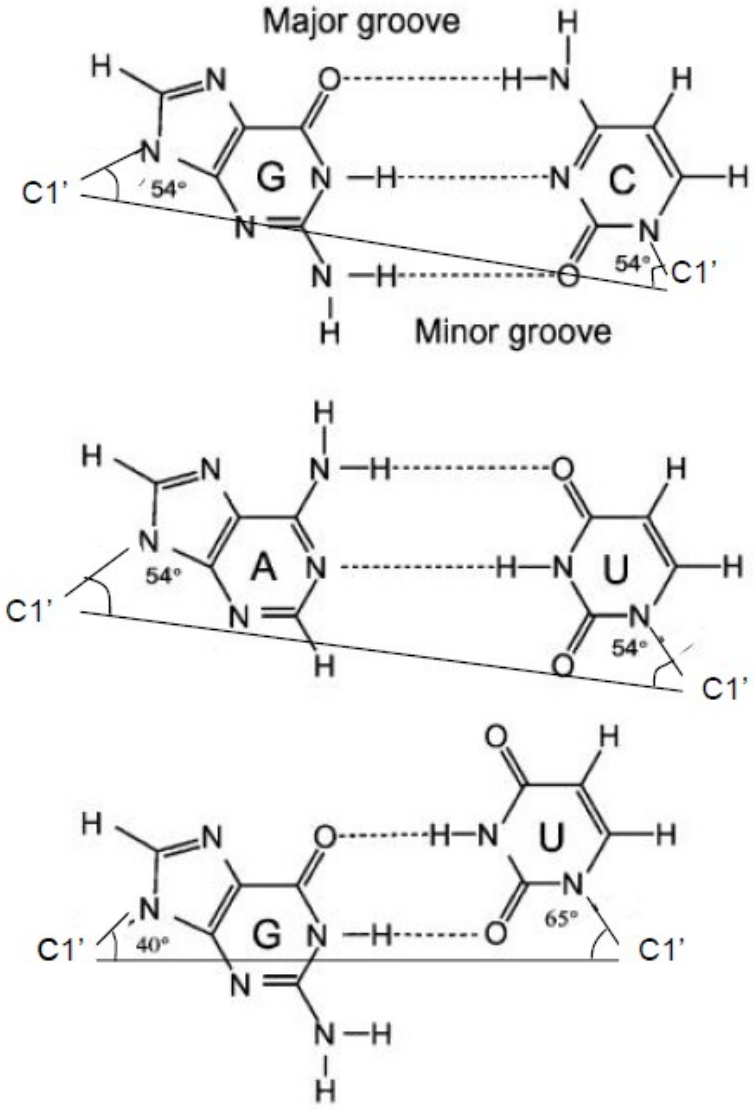


Figure 5 Watson–Crick G-C and A-U base pairs with a similar angle of ~54. The G-U “wobble base pair” is also shown (Varani and McClain 2000; Fredrick Sijenyi, Pirro Saro et al. 2012). While the G-U wobble pair is somewhat different in shape.

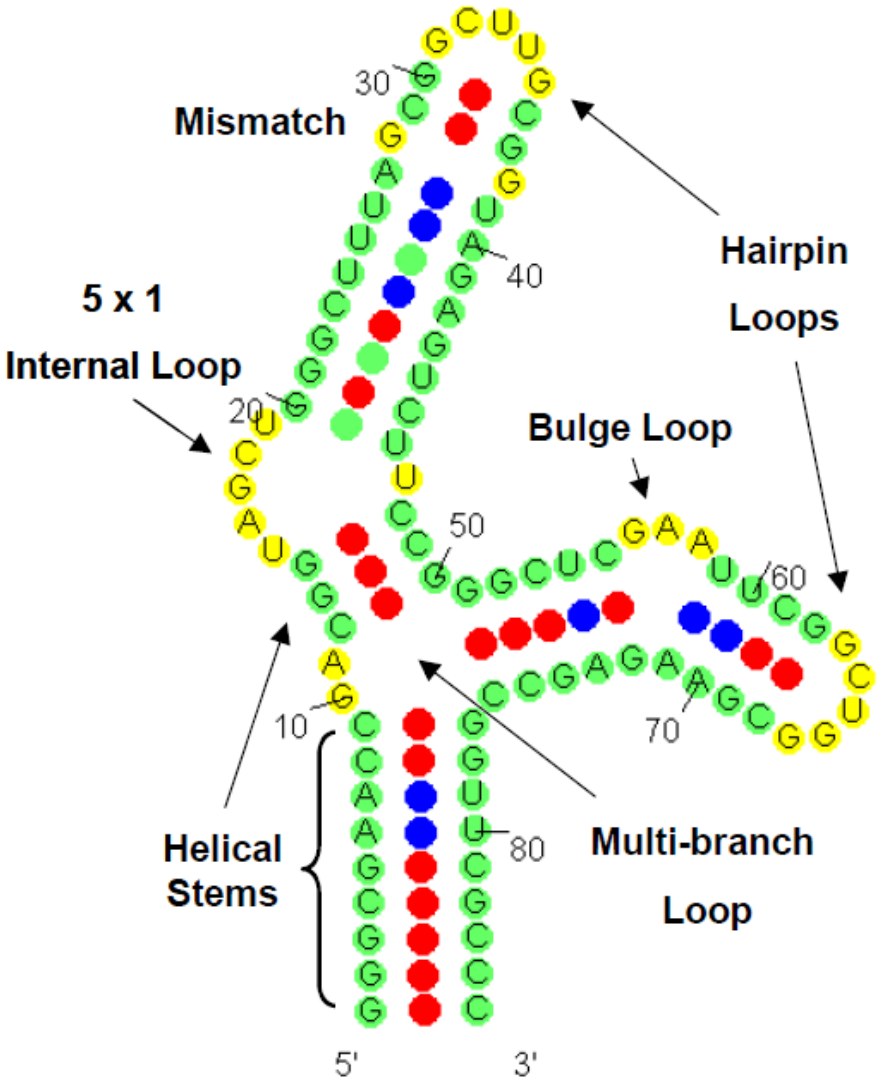


Figure 6 RNA secondary structure motifs showing representation of Watson crick base paring (Fredrick Sijenyi, Pirro Saro et al. 2012).

The 18S rRNA is composed of a region homologous to the prokaryotic 16S rRNA with several eukaryotic specific ESs (Jammie J Cannone 2002). The ESs helical elements in eukaryotes display variable lengths; however their architecture is found to be preserved (Julius Rabl, Marc Leibundgut et al. 2011). Both segments of ES3 are located on the 5' domain, ES6 and ES7 do form insertions in the central domain, and ES9 and ES12 are found in the 3' major and 3'minor domain of the 18S RNA, respectively (Julius Rabl, Marc Leibundgut et al. 2011).

Eukaryotic 18S rRNA ES6, is the longest consisting of ~250 nucleotides which form five helices that replace the bacterial helix 21. ES6 is inserted between h20 and h22 in the lower back region of the 40S body. A, B, C, and D helices form a large portion of the back of the 40S body and their structure differs considerably from a previous model (Derek J. Taylor 2009). C and D Helices are located at an equivalent position to bacterial h21. They stretch across the back of 40S and are buried underneath helix A.ES6B apical loop region is exposed and disordered in the structure, which makes it prone to chemical modification and cleavage by nucleases, as previously observed (Nygard 2004). The loop region ES3B forms a base pair with helix ES6E loop region which yields to an extended helix which leads to projection from the center of the back toward the left foot of the 40S. Earlier demonstration by computational and biochemical experiments showed this rather unusual quaternary interactions (Nygard 2004). The left foot along with ES6E is formed when the ES3B apical region packs against ES3A. ES3 and ES6 quaternary eukaryotic interaction, together with several proteins form a new domain, responsible for more prominent left foot features of the 40S and a broader back. Situated directly below the bleak is helix h16, which is shifted relative to the position in bacteria by as much as 40 Å.

In the 40S crystal structure the position of h16 is consistence with those observed in the solution by Cryo-EM of the empty yeast 40S, the 40S-eIF1 or40S-eIF1A complex (Lori A. Passmore 2007), and the canine and *Thermomyces lanuginosus* 80S (Preethi Chandramouli 2008; Derek J. Taylor 2009). Formation of a connection between the head and the body of the 40S subunit involves h16 upon binding of the initiation factors eIF1 and eIF1A, which might point to a role of this helix in initiation (Lori A. Passmore 2007).

2.4 RNA as a drug Target

The pharmaceutical industries together with the researches have always forecast their efforts more on protein, rather than the nucleic acids as suitable targets of drugs. But the acceleration and advances of studies on the RNA synthesis, structure determination and therapeutic target identification has blown open the question of using RNA as a drug target a very genuine area (Pearson and Prescott 1997). In this age due to the wealth of the three dimensional structure of the RNA in various repositories, it is possible they could be used for drug design by observing the target design. It has been a difficult task in the past to obtain a three dimensional structure for large RNAs using the crystal and magnetic resonance experimental techniques, which was just restricted to very small fragments (Pearson and Prescott 1997; Hermann and Westhof 1998; Vicens and Westhof 2003) but newer techniques and more input to the older experimental methods are leading to accurate much larger structures. The RNA on its chemical basis does not show a promising drug target in that it's made up of four different planer bases and negatively charged nucleus (Hermann 2002; Vicens and Westhof 2003). But upon the RNA adopting its conformational architecture, that shows the presence of cavities and pockets which could bind to shape specific rather than sequence specific molecules (Hermann 2002;

Sadee, Wang et al. 2011; Decher, Netter et al. 2013). The RNA cavities compels the phosphate groups to be in close proximity, that lead to an intensified importance of tightly bound water molecules, electrostatic forces and ions, magnesium divalent ions in particular that can be partly dehydrated (Vicens and Westhof 2003; Velagapudi, Seedhouse et al. 2010). Presence of non-Watson Crick pairs and bulged residues leads the formation of pockets and enlarged grooves that infer function through assembly of existing RNA motifs which are architecturally clearly diverse (Batey, Rambo et al. 1999; Vicens and Westhof 2003). The secondary structure of the RNA consisting of regular double stranded helices gives it most of its energy content and as proteins it fold to subsequent three dimensional structure free energy content of between 5-10kcalmol (Gale, Cundliffe et al. 1981; Cundliffe 1987). This gives a Nano molar range binding constant which can be achieved for a small molecule usable to compete with the final step of RNA folding (Vicens and Westhof 2003; Thomas and Hergenrother 2008). From the discussed above projections with a number of ways of inserting non Watson crick pairs within helices is very limited leading to the restriction in the number of RNA motifs that are picked as drug target (Vicens and Westhof 2003; Leontis, Lescoute et al. 2006). RNA motifs appear as decreasing size placed one inside the other, with the smaller motif associated with the larger motif (Vicens and Westhof 2003; Grabow, Zhuang et al. 2013). It has been shown that some antibiotics bind structurally in different regions of RNA molecules in the ribosome, such as streptomycin binding to the shallow groove, hygromycin binds to the deep groove of the helix, aminoglycosides at three adenine bulges and macrolides in the tunnel of the nascent polypeptide chain (Brodersen, Clemons Jr et al. 2000; Carter, Clemons et al. 2000; Hansen, Moore et al. 2003; Vicens

and Westhof 2003; Hansen, Ippolito et al. 2004; Grabow, Zhuang et al. 2013). Streptomycin interacts with only phosphate groups of many RNA (Carter, Clemons et al. 2000; Shi, Houston et al. 2011). In conclusion there are various other ways that ligands interact with the RNA that are not known yet and the growth of a rich three dimensional structural of both RNA and ligands leads to newer technique of drug discovery that may include binding and docking experiments of the structures to obtain newer and stronger intervenes.

2.5 40S Ribosomal proteins

Proteins are constituents of one or more chains of amino acids that form large biological molecules. Proteins perform various functions in cells depending on their differences in sequences dictated by the nucleotide sequence of their genes and usually result in folding into a three dimensional structure. This determines its activity that include catalytic metabolic reactions, replication of DNA, response to stimuli and transport of molecules from one location to another (Albert L. Lehninger, David L. Nelson et al. 2005). A peptide is a single linear polymer chain of amino acids bonded together by bonds that occur between the carboxyl and amino groups of adjacent amino acids residues (Albert L. Lehninger, David L. Nelson et al. 2005). The amino acid sequence of the protein is defined by the sequence of gene encoded in the specific genetic code with standard 20 amino acids, but in certain organisms such as archaea have pyrrolysine and others have selenocycteine (Yuan, O'Donoghue et al. 2010).

The unique amino acid sequence in the protein is referred to as the primary structure of the protein which in turn coils or bends into sheets to form the alpha helix or a beta pleated sheet which can exist separately or joined protein referred to as the secondary

structure. A more stable protein structure forms by further folding back upon itself and it's held together by hydrogen bonds and disulfide bridges forming a three dimensional structure also referred to as the tertiary structure. Formation of a complex structure by the interactions of two or more polypeptide chains forms the quaternary structure (Albert L. Lehninger, David L. Nelson et al. 2005) (Figure 7).

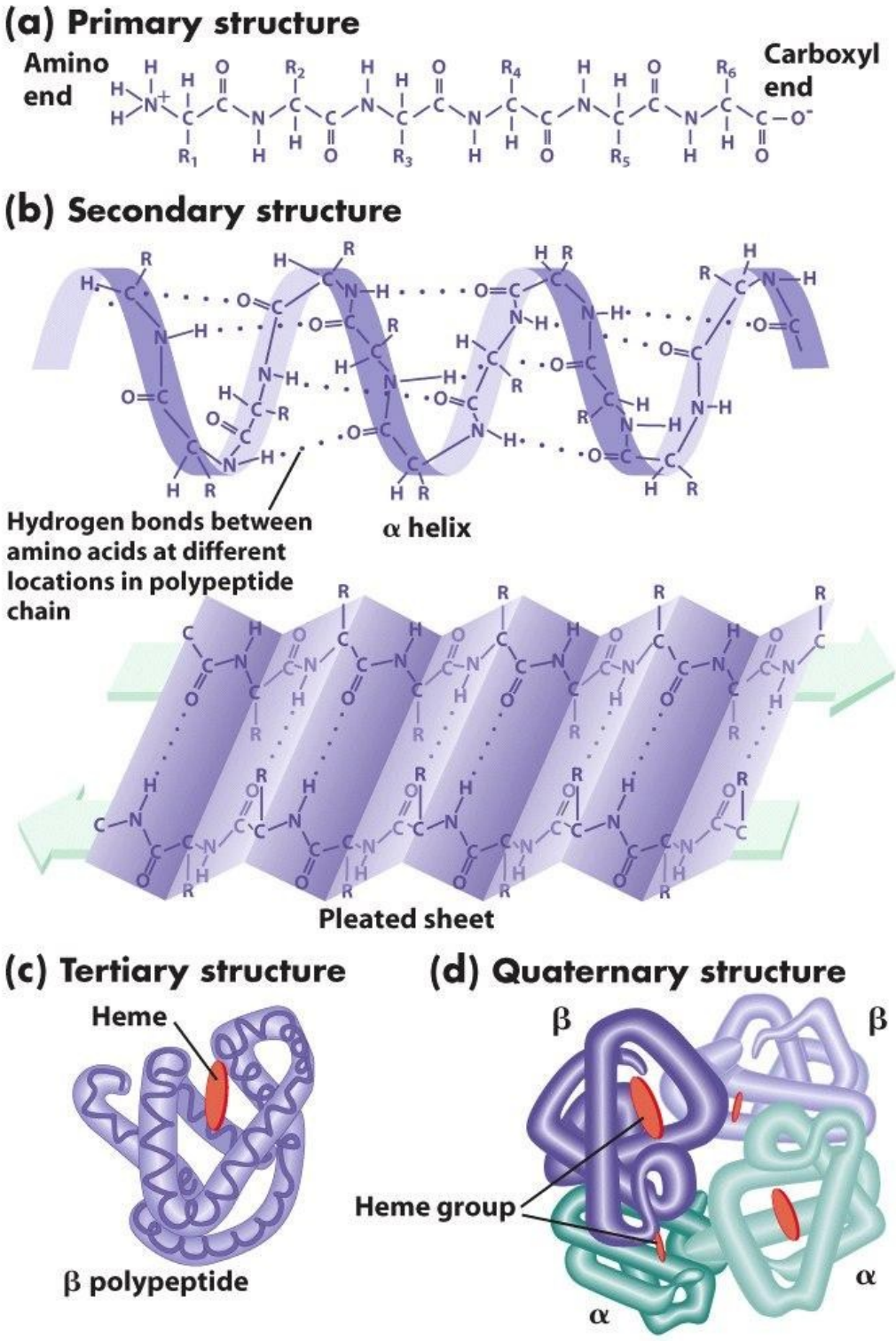


Figure 7 Protein structure hierarchy (LaMorte 2012).

Residues are often chemically modified by post translational changes during or immediately after process of protein synthesis. This alters the physical and chemical properties, folding stability, activity, finally the protein function and at times formation of prosthetic groups or co factors from attachment of proteins having non peptide group. (Vipsita, Shee et al. 2010). Stable protein complexes are often formed when two or more polypeptide chains come together and may work with other complexes to perform particular functions (Albert L. Lehninger, David L. Nelson et al. 2005).

In many cases, almost all known proteins have similar structures with other proteins and share common evolutionary origin (Murzin, Brenner et al. 1995). Knowledge of these relationships is central to understanding the structure and evolution of proteins making it an important contribution to areas of science especially in molecular biology. Proteins are basically classified using their evolutionally unit known as protein domain, which can be defined as an evolutionally unit in the sense that it's either observed in nature, or in different multi domains proteins in more than one context (Andreeva, Howorth et al. 2008). Structural classification of proteins using their protein domains are grouped according to their species and hierarchically classified into families, superfamilies, folds and classes, which embody the evolutionary and structural relationships (Murzin, Brenner et al. 1995; Conte, Brenner et al. 2002; Andreeva, Howorth et al. 2008). ***Protein species*** represents individual proteins sequences which are naturally occurring or are artificially created alternatives; ***Protein*** represents a group of similar sequences of similar functions essentially, either originating from different biological species or are representatives of different isoforms which are within the same organism; ***Family*** represents group of proteins with sequences related however characteristically distinct functions;

Superfamily is a group of protein families with common structural and functional features shown to be from a common ancestor evolutionally; ***Folds*** groups are purely structural classified similar superfamilies with different characteristic features and finally ***Class*** are further classification of folds based mainly on their secondary structure organization and content (Murzin, Brenner et al. 1995; Conte, Brenner et al. 2002; Andreeva, Howorth et al. 2008).

The eukaryotic 40S subunit structure contains 33 proteins, 18 of which are absent in bacteria (Armache, Jarasch et al. 2010; Armache, Anger et al. 2013). Additionally, over half of the conserved ribosomal proteins contain eukaryotic-specific extensions. Three of the eukaryotic ribosomal proteins are positional analogs that have replaced bacterial 30S proteins: Bacterial rpS6p and rpS18p at the platform are replaced by rpS1e, whereas bacterial rpS16p at the foot has been substituted by rpS4e (Julius Rabl, Marc Leibundgut et al. 2011). RpS30e eukaryotic ribosomal protein occupies a similar space as bacterial rpS4p, but unlike rpS4p it extends into the inter-subunit interface (Barbara S. Schuwirth, Maria A. Borovinskaya et al. 2005). The feet of the eukaryote ribosome is extensively remodeled to the bacterial ribosome, where h9, ES3, and ES6E of the 18S rRNA and eukaryotic ribosomal proteins rpS4e, rpS6e, and rpS8e have replaced bacterial rpS20p. Eukaryotic ribosomal proteins rpS10e, rpS12e and rpS31e are bound to a reduced h33 of the 18S rRNA, giving the eukaryotic protein beak essentially the same conserved shape as observed for the bacterial 30S subunit (Keqiong Ye 2009).

The expression of rpS31e as an ubiquitin fusion protein, stretches from the A site close to the decoding center towards the beak, where there is the location of zinc finger motif (Kent L. Redman 1989). N-terminal ubiquitin domain failure to cleave would result in a

40S subunit that cannot engage in initiation due to the steric hindrance of the decoding center by attached domain. In fact, mutations preventing cleavage of the ubiquitin moiety from rpS31e are lethal in yeast (Thierry Lacombe 2009). Additional eukaryotic proteins are mostly located at the back of the small subunit, contact the 18S rRNA, bind to each other or interact with the other conserved proteins. On the contrary, extensions specific to the eukaryotic conserved proteins are found clustering and interacting primarily with eukaryotic specific proteins. In prokaryotes, five of 20 ribosomal proteins interact with no other ribosomal protein (Julius Rabl, Marc Leibundgut et al. 2011). The difference between the eukaryotic from prokaryotic found that all ribosomal proteins *of T. thermophila* 40S subunit contact at least one other ribosomal protein, and the total interaction surface between ribosomal proteins in the 40S subunit is almost four times the 30S ribosomal protein-protein interaction surface (40S, ~25,000 Å2; 30S, ~6500 Å2) (Julius Rabl, Marc Leibundgut et al. 2011). Although the increased number of ribosomal proteins in the 40S is partly responsible for the larger interaction surface, the average interaction surface of a eukaryotic protein is nonetheless larger than a bacterial protein in the 30S subunit by a factor of 2.4 (Adam B en-Shem 2010; Ben-Shem, Garreau de Loubresse et al. 2011). Long extended C- or N-terminal regions of proteins, frequently involved in RNA interactions in bacteria, are occasionally found to be responsible for large distance-bridging contacts between eukaryotic proteins in the 40S subunit. For example, the C terminus of rpS3e stretches from the core of rpS3e near the beak toward RACK1, while the C terminus of rpS17e meanders away from RACK1 along the back of the ribosome over a distance of at least 48 Å, where it binds to rpS0e (Julius Rabl, Marc Leibundgut et al. 2011). The 40S subunit positioning of the ribosomal proteins allows

rationalization of biochemical and genetic data on the ribosome biogenesis and human disease such as 5q− syndrome and Diamond-Black fan anemia, although distribution of this ribosomal proteins shows a striking correlation with biochemical and genetic data regarding their involvement in different stages of ribosome biogenesis (Burwick, Shimamura et al. 2011). Ribosomal protein found scattered the 40S subunit are associated with human disorder, the later suggesting that these mutations disturb ribosome stability or assembly rather than affecting specific function of the 40S subunit (Scheper, van der Knaap et al. 2007; Burwick, Shimamura et al. 2011).

2.6 Protein homology modeling

Homology modeling for evolutionally related proteins is a process that involves two proteins that have a close common ancestor. An alignment score can be obtained to show similarities between both organism's sequences, with a lower alignment score often suggesting different structural and functional moieties (M.Lesk 1986). Alignments of both the query and template sequences structures are in turn used to run homology modeling. Inaccuracies of homology models are often traced to the lower sequence identity between the two (Margelevicˇius 2005). Either way, models that are created by this technique are usually analyzed thoroughly for structural consistency such as correct bond lengths, angles and residue steric clashes. Modeling of proteins in this thesis was performed primarily using the Swiss Model (Peitsch 1996; Arnold, Bordoli et al. 2006). This is an online server that allows users to enter their preferred template and enter in their sequences that are to be modeled and does the modeling, returning the created models as outputs.

2.7 RNA homology modeling

RNA modeling is more challenging than protein modeling due to structural complexities associated with RNA. For instance the RNA backbone has six dihedral angles while proteins have only two. In additions RNA residues are substantially larger in size than the protein residues. Though RNA prediction field has advanced, more work still needs to be done on developing conformational sampling and searching methodologies that have the ability of locating the minimal energy in a tractable time period on a personal computer. (Das and Barker 2008).Various research groups have worked to increase the size limit and the accuracy of predicted RNA structures. The Santa Lucia group has developed the software package *RNA123,* which is a suite of tools for homology modeling, *de novo* structure prediction and analysis of the RNA structures (Fredrick Sijenyi, Pirro Saro et al. 2012). This tool has demonstrated the ability to predict structures as large as the bacterial ribosome. The modeling process in *RNA123* involves employing a series of algorithms that account for substitutions, insertions, and deletion and gap closing and performing a final energy minimization to finally determine a low energy structure. *RNA 123* was the homology and de novo modeling tool used in the reported study.

Chapter Three

3.0 Methodology

Most homology-modeling methods consist of four steps: Template selection, Target template alignment, Model building, and evaluation of the model produced. Modeling of a structure is achieved by iteratively repeating these steps until the desired model is produced. A number of different techniques for model building have been developed over a period of time to suite the growing demand of good structures (Tom L. Blundell 1987; Nicholas Guex 1997; O. Lund 1997). For example, the approach adopted by SWISS-MODEL server can be described as rigid fragment assembly which was first implemented in Composer (Tom L. Blundell 1987), and is outlined below.

3.1. Template selection

The Template was selected for given protein or RNA where the sequences of the template structure library are searched (Stephen F. Altschul 1990; Stephen F. Altschul 1997). If these templates covered a distinct region of the target sequence they were scored to give a homology score which in turn was used to select the best pair. Further information useful for template selection was gathered and added to the file header, e.g. probable quaternary structure (Schwede, Kopp et al. 2003), quality indicators like empirical force field energy (Walter R. P. Scott 1999) or ANOLEA mean force potential scores (Patterson 1998) for SWISS-MODEL server. Most models required that the template be in the PDB format in order to allow for a stable and automated work flow of the server. SWISS-MODEL picked the coordinate files of individual proteins chains and for the unreliable entries e.g.

theoretical models and low quality structures providing only Ca coordinates, are removed. In this study, coordinates files for every individual protein (~30) in the 40S subunit were generated. This was done by simply opening the PDB files of the template used in this study (*T. Thermophila*) on a text editor and copying coordinates corresponding to each chain for all proteins. A more detailed understanding of the template was done by referencing the PDB ID of the template in the protein databank website (http://www.rscb.org/pdb/explore.do). The crystal structure however did not contain all the residues known to belong to each protein based on the gene bank derived sequences. This was however not an issue as the missing sections were *de novo* modeled automatically.

3.1.1 Alignment

After selecting a reasonable template or in case there was more than one template the best was selected by superimposing them using an iterative least squares algorithm and obtaining the best. In SWISS _MODEL, generation of a structure alignment was done after removing incompatible templates, i.e. omitting structures with high Ca root mean square deviations to the first template. A local pair-wise alignment of the target sequence to the main template structures was calculated (24), followed by a heuristic step to improve the alignment for modeling purposes. The placement of insertions and deletions was optimized considering the template structure context. In particular, isolated residues in the alignment ('islands') were moved to the flanks to facilitate the loop building process. This alignment process was similar in both RNA and Protein alignments (Fredrick Sijenyi, Pirro Saro et al. 2012). In this study, the template coordinates of the

~30 proteins were obtained from the *T. Thermophila* crystal structure PDB ID 2XZM, the query sequences from the *P. falciparum* proteins were confirmed using a BLAST search (http://www.ncbi.nlm.nih.gov/protein) from the *P. falciparum* genome. Complete sequences of our target proteins were obtained and saved in a text file for each protein.

3.1.2Model building

Generation of the model core begun by averaging the backbone atom positions of the template structure. This was followed by weighting template by their sequence similarity to the target sequence, while significantly deviating atom positions were excluded. Coordinates of the template cannot be used to model regions of insertions or deletions in the target-template alignment. To generate those parts, an ensemble of fragments compatible with the neighboring stems was constructed using constraint space programming (CSP) for proteins or using a *de novo* algorithm in *RNA123*. This was done using a scoring scheme where the best loop was selected, based on favorable force field energy, steric hindrance and favorable interactions like hydrogen bond formation. If no suitable loop could be identified, the flanking residues are included to the rebuilt fragment to allow for more flexibility. In cases where CSP did not give a satisfying solution and for loops above 10 residues, a loop library derived from experimental structures was searched to find compatible loop fragments (Simon C. Lovell 2000).

3.1.3 Side chain modeling

The model side chains reconstruction was based on the weighted positions of corresponding residues in the template structures. The model side chains were built by iso-sterically replacing template structure side chains starting with conserved residues. Selection of possible side chain conformations from a backbone dependent rotamer library, which has been constructed carefully taking into account the quality of the source structures was performed (Simon C. Lovell 2000). A scoring function assessing favorable interactions (hydrogen bonds, disulfide bridges among others depending on whether RNA or proteins) and unfavorably close contacts was applied to select the most likely conformation. In rRNA modeling, depending on the complexity of the rRNA, homology modeling was done in parts by dividing the structure into the different ribosomal domains;5'major, central, 3'minor and 3'major domains. This allowed independent domain modeling which was a better approach than trying to model the entire 18S rRNA as it allows for easier manipulation of the alignment which was crucial to generating a viable structure.

3.1.4 Energy minimization

Deviations in the protein or RNA structure geometry, which had been introduced by the modeling algorithm when joining rigid fragments were regularized in the last modeling step by steepest descent energy minimization using the GROMOS96 force field (Walter R. P. Scott 1999) for SWISS-MODEL/. For *RNA 123*, empirical force fields were used to

detect parts of the model with conformational errors. Energy minimization or molecular dynamics methods were in general not able to improve the accuracy of the models, and were used in the modeling process to regularize the structure (Moult 1996; Levitt 1999). However, the successful application of restricted molecular dynamics for improving homology models has recently been reported for a few test cases (J.A. Flohil 2002). To derive more general rules of engagement of molecular dynamics, further systematic experiments had to be conducted. The four modeling steps—template superposition, target template alignment, model building and energy minimization— having been implemented in the program ProModII in ANSI C (Nicholas Guex 1997). For The RNA *RNA123* had integrated most of these algorithms to run in three simple steps, preprocessing, alignment and finally modeling (Fredrick Sijenyi, Pirro Saro et al. 2012).

3.1.5 Structure Validation

Once modeling was performed, validation of the generated models was carried out using several validation tools such as MOL-PROBITY (Chen, Arendall et al. 2010), a free online tool available at (http://kinemage.biochem.duke.edu/) or Discovery studio Accerlys software (Studio 2009). This helped in the addition of hydrogen's to the crystal structure. Other tools employed in this process included PRO-CHECK or MATCH-CHECK (Laskowski, Rullmannn et al. 1996), these too were available as free online tools and were valuable in structure validation.

3.2 Proof of concept

There are majorly two main ways of obtaining the 3D structure of the eukaryotic ribosome these ways can be either theoretical or practical methods. In the practical or

experimental methods, we have X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy. In X-ray crystallography the atomic or molecular structure of a crystal is obtained by measuring the diffracted X-ray beam initially projected to the query molecule. X-rays diffract into many angles which can be measured together with their intensities to produce a three dimensional structure of the electron densities within the crystal or molecule (Massa 2011; Barty, Küpper et al. 2013). From this information, positions of atoms can be determined together with their chemical bonds among other information. The major shortfall of this method is that for more complex structure it's not easy to crystalize and also the three dimensional structure obtained become less well resolved for a given number of observed reflections. In nuclear magnetic resonance, a magnetic nuclei absorbs and re-emits an electromagnetic radiation in a magnetic field. The energy that is emitted at a specific resonance frequency depending on the strength of the magnetic field and its magnetic properties of the isotope of the atom allow the observation of specific quantum mechanical magnetic properties of the atomic nucleus (Andrew 2009). In Cryo-electron microscopy (cryo-EM), a sample is studied at cryogenic temperature (generally liquid nitrogen) to develop a biological structure a method often known as transmission of electron microscopy. Specimens are observed as they are found in their native environment, in contrast to X-ray crystallography where samples are placed in a non-physiological environment which may lead to functionally irrelevant conformation changes. Limitation of this method is that the low resolution maps are not enough to allow for unambiguous model generation on the basis of electron microscopy only, and structures obtained need others generated by

Crystallography to be interpreted (Frederik, Stuart et al. 2011; Milne, Borgnia et al. 2013).

The theoretical methods involve comparative (Homology) modeling, threading and *Ab initio* method such as *De novo*. Homology modeling is a good technique to obtain biological three dimensional structures where a good template is available in comparison with other methods due to time it takes to generate valid biological structures. *De novo* methods are best and are applied where homology model cannot be used due to lack of a good template.

Homology or comparative modeling of biological molecules is the construction of atomic resolution models of the query molecule from its amino acid or nucleotide sequence and an experimental three dimensional structure of a related homologous molecule known as the template. In this case, homology modeling relies on the identification of known structures likely resembling the structure of the query sequence. Homology modeling in *RNA123* is illustrated in Figure 8 . For a proof of concept project, we used homology modeling to generated structures of a known crystalline structure of the 40S subunit of *Saccharomyces cerevisiae* (RSCB ID 3O30) from its homologous eukaryotic structure - the Crystal Structure of the Eukaryotic 40S Ribosomal Subunit from *Tetrahymena thermophila* (RSCB ID 2XZM). *Saccharomyces cerevisiae* nucleotide sequence of the 18S rRNA and ~34 protein amino acid sequences of the 40S subunit was used as the query with the 40S subunit of *Tetrahymena thermophila*(2XZM RSCB ID) being used as the template. With negation of any other information in the *Saccharomyces cerevisiae* other than the sequence information the homology modeling methodology

was followed to obtain the model structure. The models generated using this method were then compared to the actual crystal structture in order to asses the quality of the models. Comparisons were made by superimposing the models to the crystal structure and calculating the RMSDs (Figure 9). Based on the obtained results, we concluded that homology modeling technique was reliable but dependend on the similarity between the two organisms. Therefore this method can be used to generate reliable structures fairly quickly which can hasten the structural studies of organisms whose structure do not currently exist.

3.3 Homology /comparative modeling process

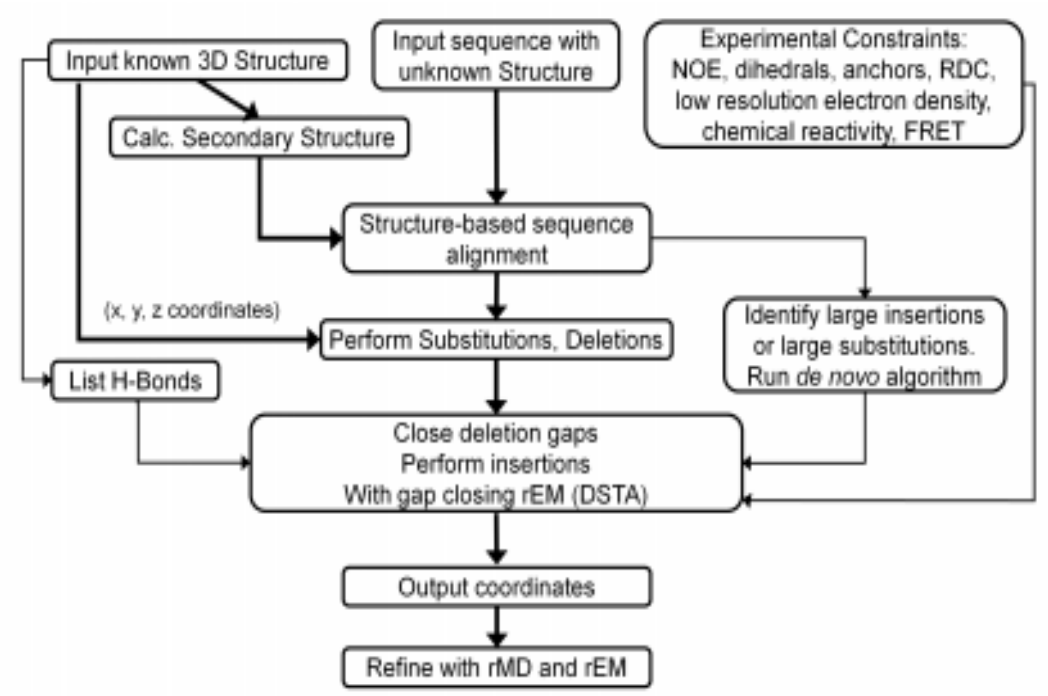
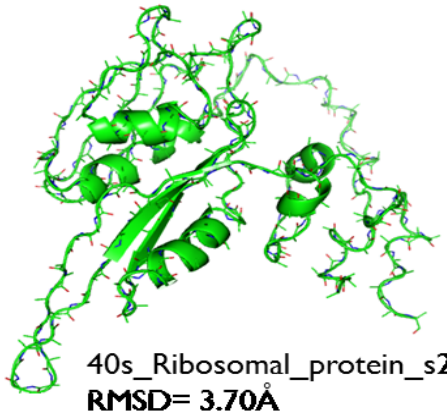


Figure 8 Schematic diagram showing general systematic Homology /comparative modeling process.

3.4 Preliminary results

The results that were obtained showed that the homology modeling process would be an ideal theoretical approach that could be used to predict the three dimensional structure of molecules that are unknown giving more biological insight on their functional role in the system. Some of the results are shown below in figure

(A)



(B)

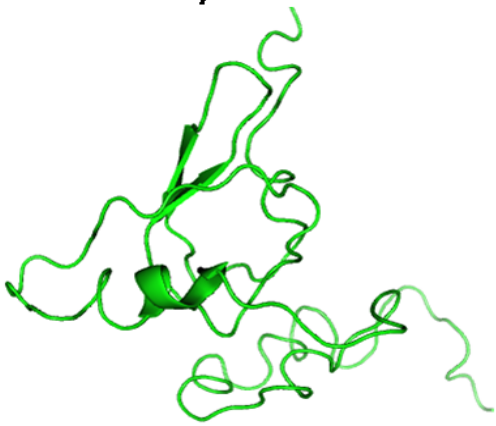


Figure 9 A and B Three Dimensional structures of some of the modeled proteins of the *Saccharomyces cerevisiae* with their RMSD compared to the crystalline models.

Saccharomyces cerevisiae 40S Protein and RNA Models	**RMSD**
40S ribosomal protein S15	3.72Å
40S ribosomal protein S16	2.94Å
40S ribosomal protein S17-A	3.40Å
40S ribosomal protein S18	4.12Å
40S ribosomal protein S20	3.83Å
40S ribosomal protein S23	3.78Å
40S ribosomal protein S3	2.97Å
40S ribosomal protein S2	3.70Å
40S ribosomal protein S5	2.40Å
rRNA	
3'Minor domain	3.98Å
3'Major domain	4.24Å

Table 1 *Saccharomyces cerevisiae* 40S ribosomal subunit showing some proteins modeled with their RMSD.

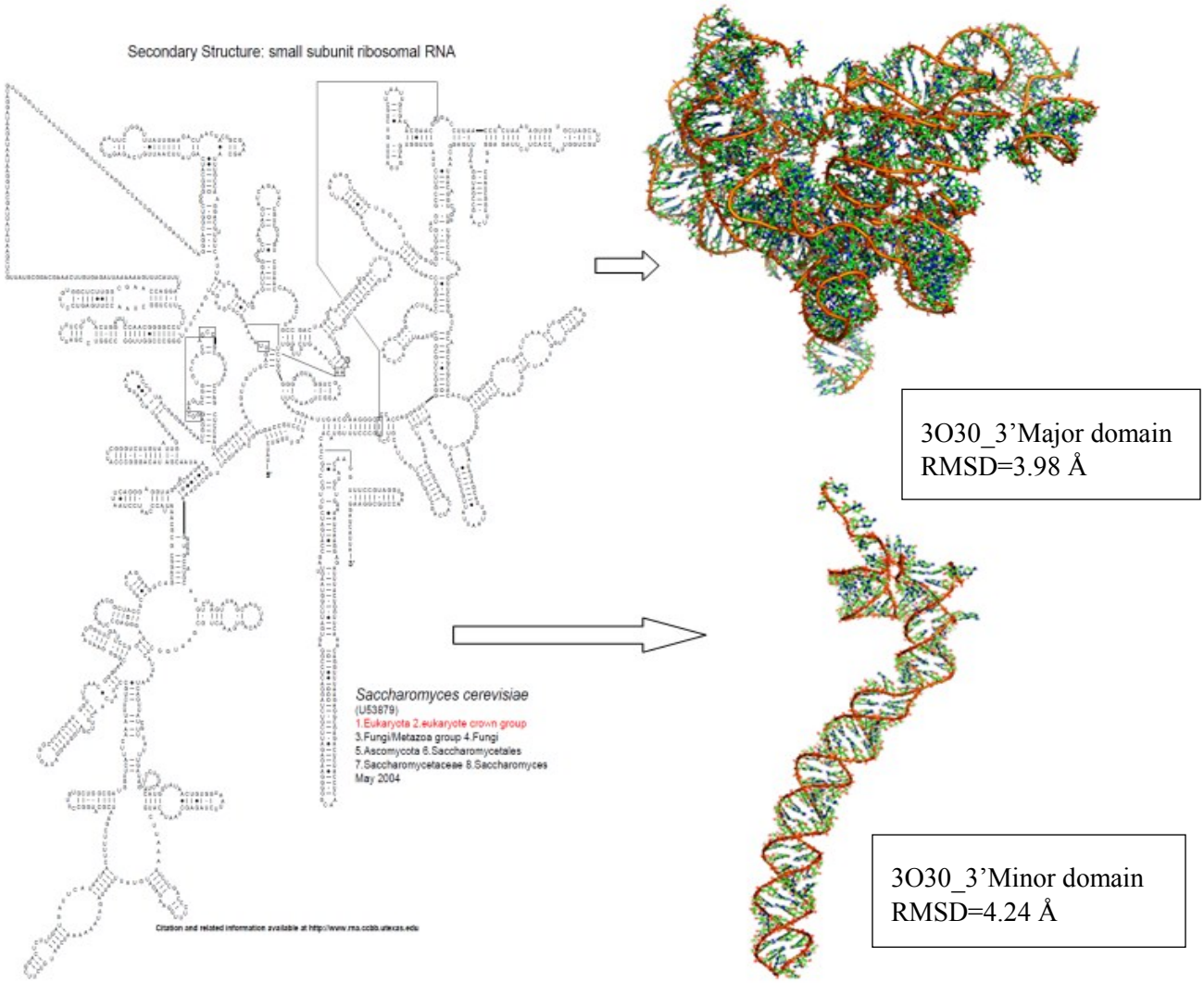


Figure 10 Diagram showing the two dimensional structure of *Saccharomyces cerevisiae* with modeled domains 3'Major and 3'Minor with their RMSDs

3.5 Conclusion of the Preliminary results

From the obtained results in this analysis it was shown that theoretical methods such as the homology modeling with the above methodology could be used to model biological structures without known three dimensional structures that would be quite necessary to help understand their role in life in this era of computation power.

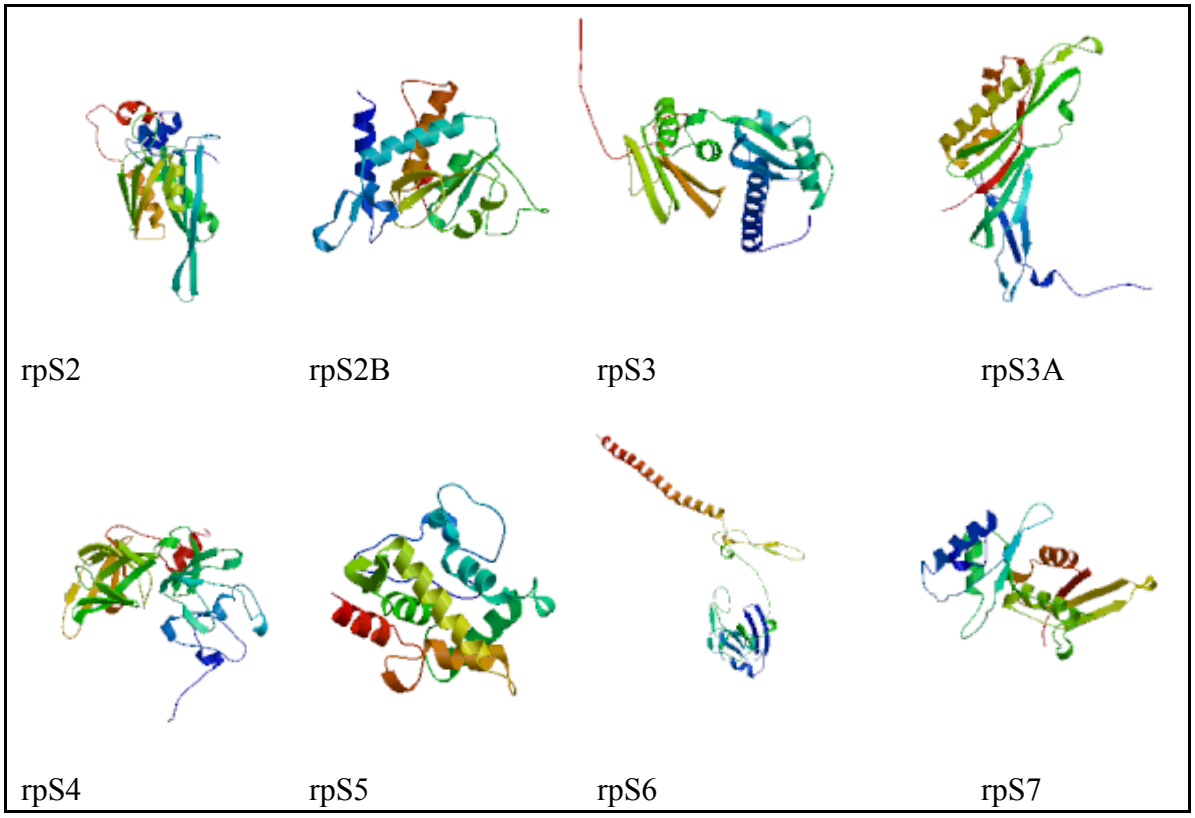
Chapter Four

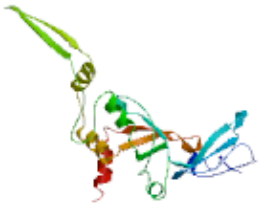
4.0 Results

4.1 Protein models of *Plasmodium falciparum*

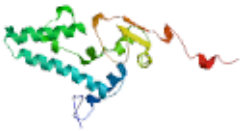
4.1.1 Plasmodium falciparum 40S proteins

Plasmodium falciparum has 34 proteins which were successfully homology modeled using the crystal structures of eukaryotic *Tetrahymena thermophila* 40S proteins which were generated from crystallization by experimental method using x-ray crystallography.



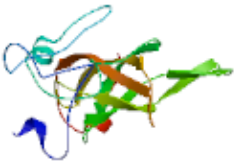


rpS8e

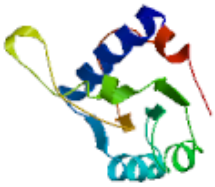


rpS9

rpS10

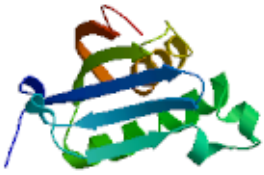


rpS11



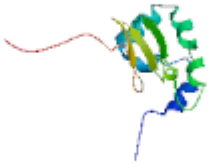
rpS12

rpS13



rpS14

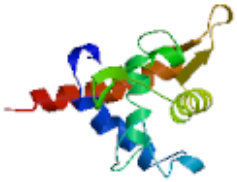
rpS15A



rpS15.19

rpS16

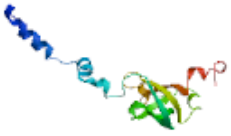
rpS18



rpS19

rpS20e

rpS21e



rpS23

rpS24

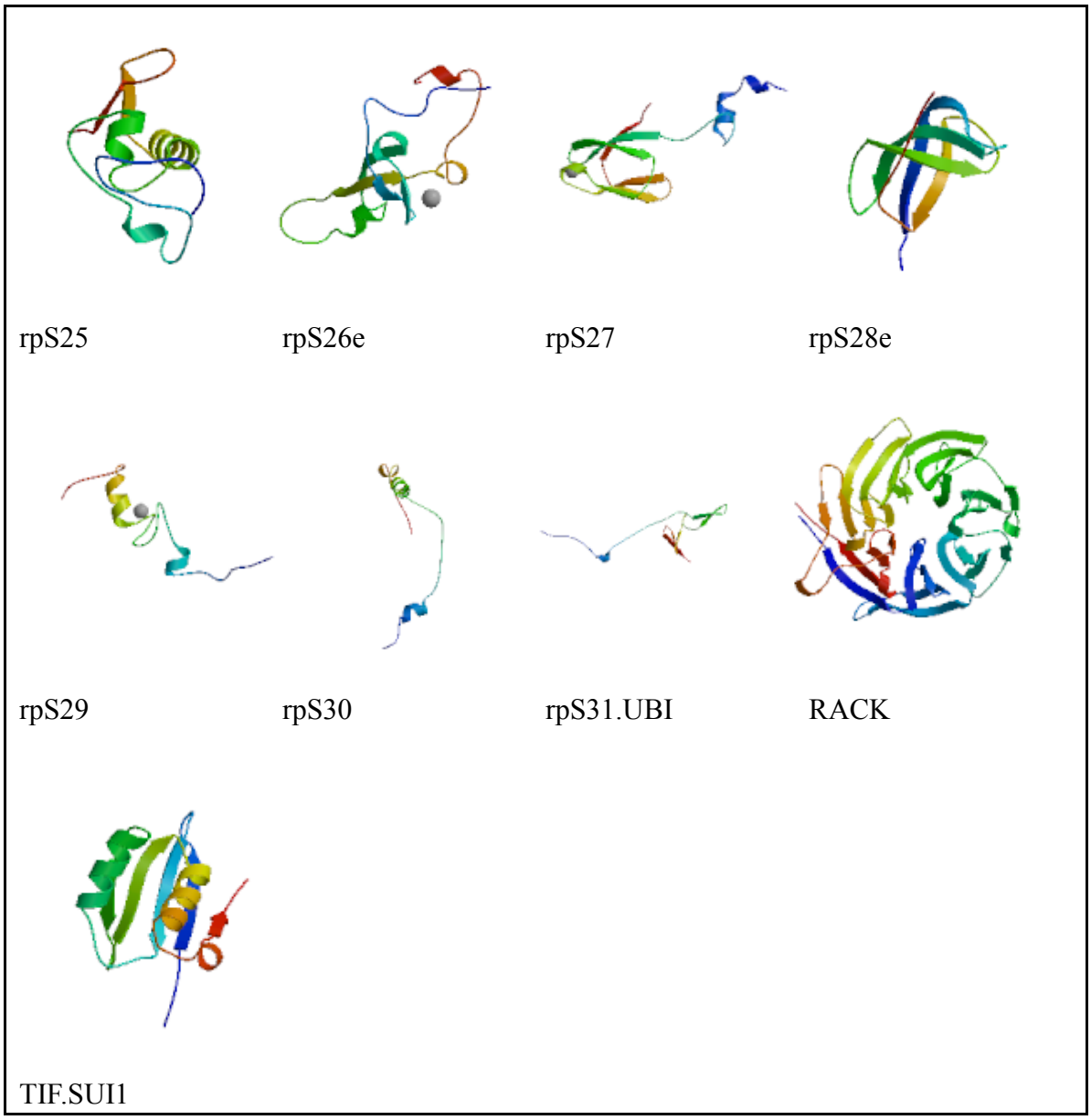


Figure 11: 3Dstructure of all ribosomal proteins of the 40S subunit of *Plasmodium falciparum*. Backbone of all the proteins as shown colored to the distribution (structures shown), and their names also indicated below.

	Plasmodium falciparum protein model	No of model residues	2XZM_*Tetrahymena thermophila* Template	No of Template residues	PDB ID	CHAIN	Pymol RMSD
1	S2	272	S5	296	Q23KG1	E	(RMSD=1.16)
2	S2B	241	S0E	241	E6PBR9	B	(RMSD=1.73)
3	S3	221	KH domain containing	243	Q22AV9	C	(RMSD=0.78)
4	S3A	265	S3A	262	Q23DE3	4	(RMSD=1.37)
5	S4	261	S4	260	P0C233	W	(RMSD=1.04)
6	S5	195	S7 Containing	200	Q22WU7	G	(RMSD=0.61)
7	S6	306	S6E	293	A4VD76	Y	(RMSD=0.80)
8	S7	194	S7E	197	E6PBS5	3	(RMSD=2.70)
9*	S8e	218	S8	208	Q22AV0	2	(RMSD=2.20)
10	S9	189	S4 Containing	181	P0C233	D	(RMSD=0.37)
11	S10	137	PLECTIN S10 domain	162	Q24F70	7	(RMSD=0.84)
12	S11	161	S17 containing	157	Q22B78	Q	(RMSD=0.66)
13	S12	141	L7AE Containing	126	Q22W26	U	(RMSD=1.46)
14	S13	151	S13E	153	E6PBT1	O	(RMSD=0.32)
15	S14	151	S14E	151	E6PBT2	K	(RMSD=0.22)
16	S15A	130	S8 Containing	130	Q238Q8	H	(RMSD=0.37)
17	S15.19	145	S15E	144	E6PBT3	S	(RMSD=0.40)
18	S16	144	S16E	145	E6PBT4	I	(RMSD=0.37)
19	S17	137	S17E	130	E6PBT5	V	(RMSD=0.40)
20	S18	156	S18E	155	E6PBT6	M	(RMSD=0.47)
21	S19	170	S19E	155	E6PBT7	T	(RMSD=1.09)
22	S20e	118	S10 containing	120	Q22DC6	J	(RMSD=0.94)
23	S21e	82	S21E	97	E6PBT9	Z	(RMSD=0.46)
24	S23	145	S12	142	P06147	L	(RMSD=0.39)
25	S24	133	S24E	149	E6PBU2	P	(RMSD=1.01)
26	S25	105	S25E	143	E6PBU3	8	(RMSD=1.61)
27	S26e	107	26E Containing	119	Q23PT4	5	(RMSD=1.08)
28	S27	82	S27E	81	Q22CK0	6	(RMSD=0.32)

29	S28e	68	S28 Containing	68	Q234G5	1	(RMSD=0.83)
30	S29	54	S29E	55	Q22MB0	N	(RMSD=0.55)
31	S30	58	S30E	80	E6PBU8	X	(RMSD=3.28)
32	S31	149	S31E	189	E6PBU9	9	(RMSD=0.36)
33	RACK	323	RACK1	343	Q24D42	R	(RMSD=3.15)
34	TIF SUI1	115	EIF1	101	E6PBV1	F	(RMSD=0.44)

Table 2: Summary of *Plasmodium falciparum* 40S ribosomal proteins models RMSDs and protein classification showing no of residues their homologous counterparts of the template (2XZM_*Tetrahymena thermophila*), their number of residues, PDB ID, and chain ID. (RMSD determined using PyMOL)

4.1.2 Protein structure Validation

Model	Query	subject	z scores	Number of equivalent residues	RMSD of C-alphas (Å)
s8	mol12	mol22	27.5	199	0.3
s9	mol1d	mol2f	26.7	179	0.1
s6	mol1y	mol2y	24.6	228	0.4
s5	mol1g	mol2g	33.2	187	0.3
s4	mol1w	mol2w	41.1	256	0.3
s7	mol13	mol23	27.9	181	0.7
s3	mol1c	mol2c	30.9	219	0.4
s3A	mol14	mol24	30.5	211	0.6
s2	mol1e	mol2e	35.7	227	0.4
s2B	mol1b	mol2b	37.3	197	0.2
s10	mol17	mol27	19.5	101	0.6
s11	mol1q	mol2q	26.3	156	0.3
s12					
s13	mol1o	mol2t	22.7	148	0.1

s14	mol1k	mol2k	24.3	122	0.1
s15a	mol1h	mol2h	27.5	129	0.1
s15	mol1s	mol2s	21.3	116	0.4
s16	mol1i	mol2i	26.1	141	0.1
s17	mol1v	mol2v	18.2	109	0.1
s18	mol1m	mol2m	24.6	154	0.1
s19	mol1t	mol2t	26.6	144	0.3
s20e	mol1j	mol2j	17.6	102	0.4
s21e	mol1z	mol2z	13.4	82	0.1
s23	mol1l	mol2l	23.2	140	0.2
s24	mol1p	mol2p	19.9	128	0.2
s25	mol18	mol2b	14.2	84	0.6
s26	mol15	mol25	14.6	96	0.5
s27	mol16	mol26	14.6	80	0.1
s28e	mol11	mol21	14.8	65	0.4
s29					
s30					
rack	mol1r	mol2r	54.2	312	1.3
tif sui1	mol1f	mol2f	20.3	87	0.1
s31	mol1x	mol29	8.9	72	0.1

Table 3: RMSD of the *Plasmodium falciparum* protein models using Procheck a validation tool. The models that was input as query and a subject template to be analyzed the z scores were obtained and the number of residues that match in shown and finally an RMSD indicated

4.2 Plasmodium falciparum 18S r RNA

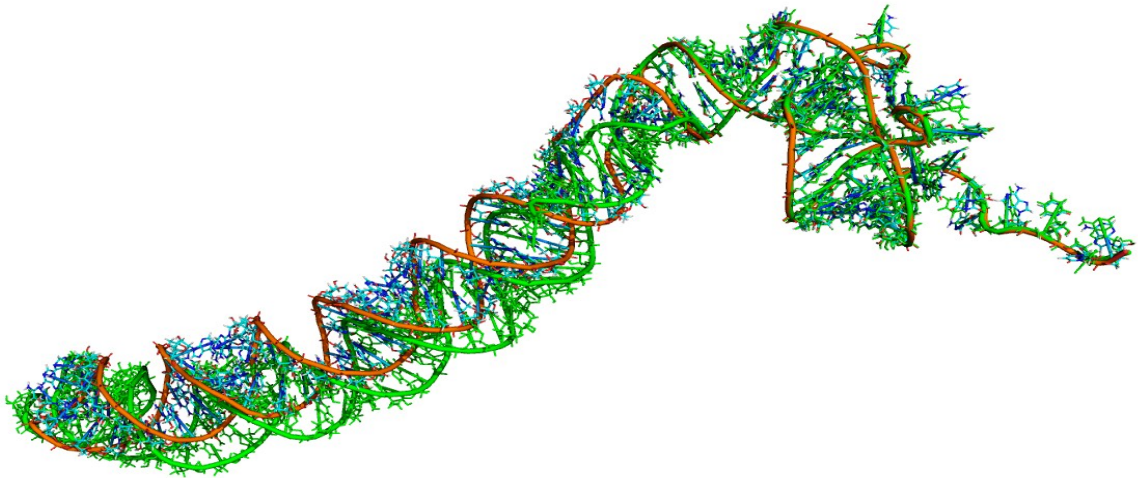


Figure 12: The modeled 3'Minor domain of *Plasmodium falciparum* superimposed with the template *Tetrahymena thermophila* 3'Minor domain.

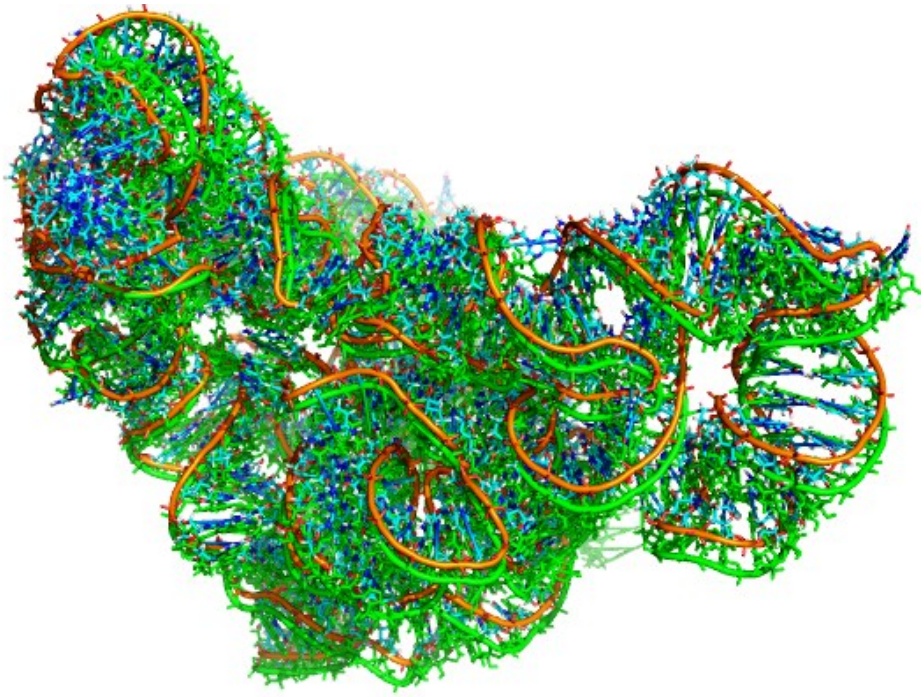


Figure 13: 3'major model of *Plasmodium falciparum* superimposed with template *T thermophila*

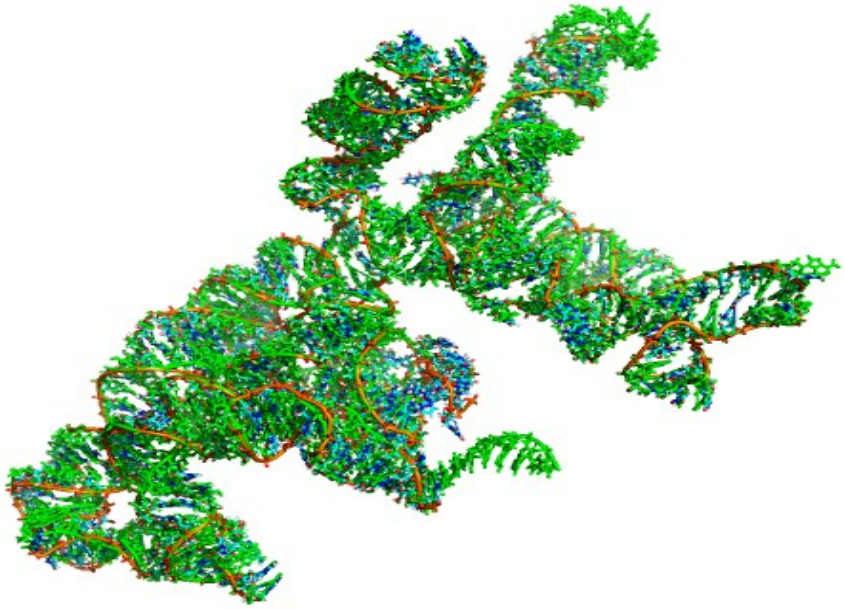


Figure 14: Modeled central domain of *Plasmodium falciparum* superimposed with the template *Tetrahymena thermophila.*

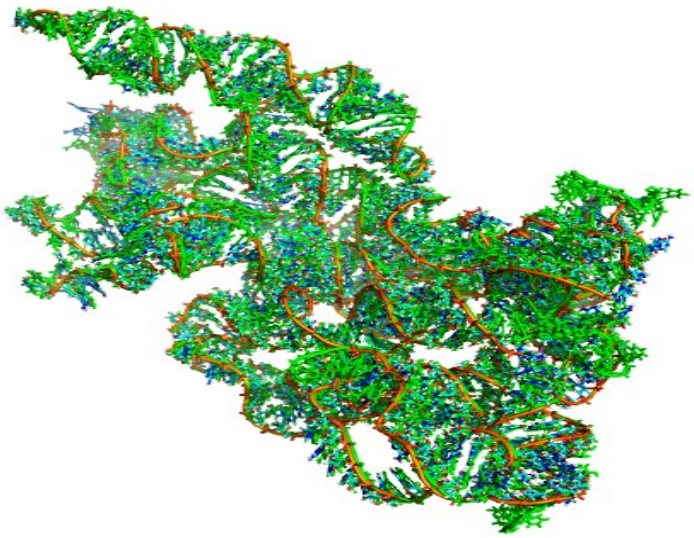


Figure 15: 5'major domain model superimposed with template

Combined structure 18S rRNA of plasmodium falciparum

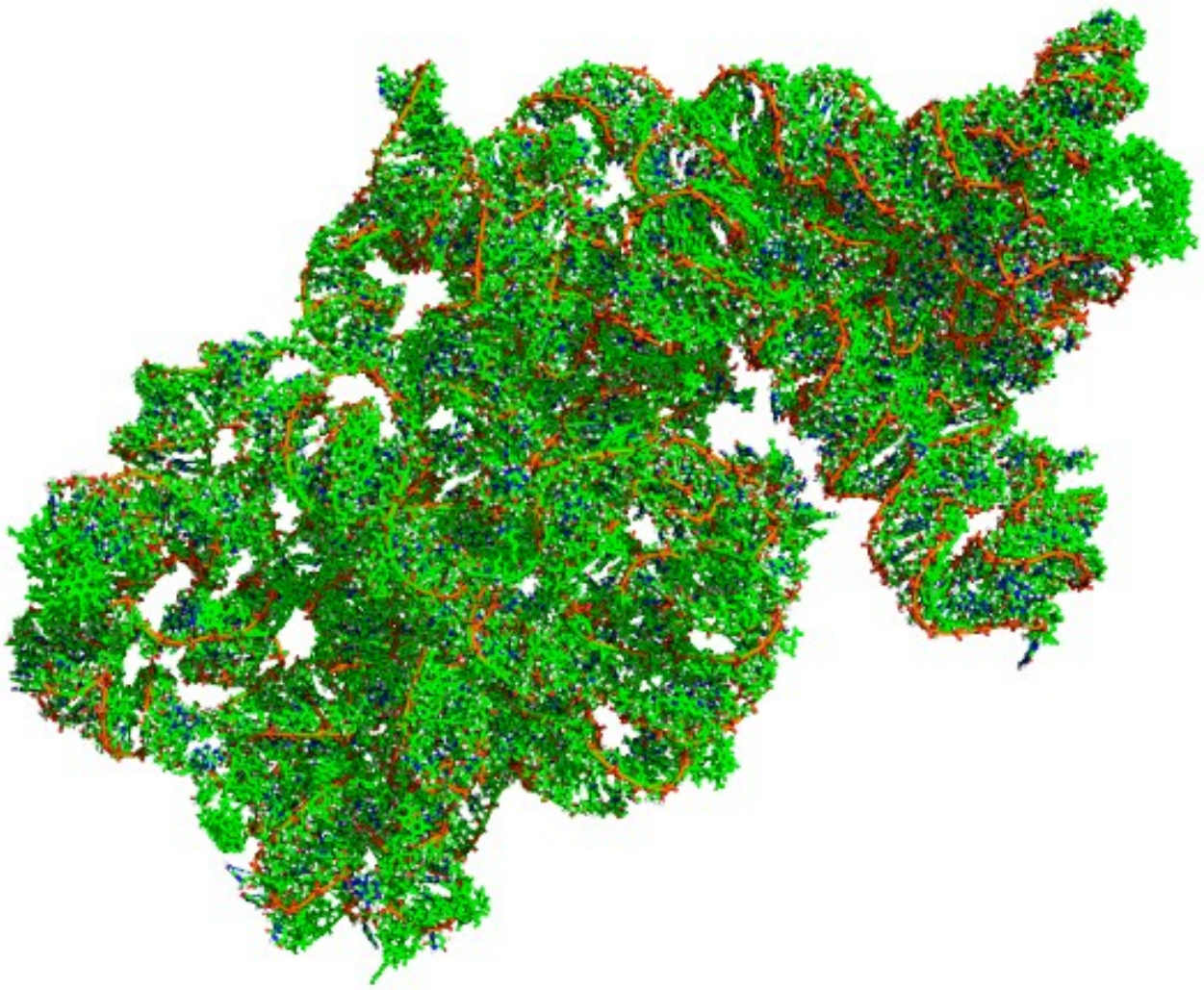


Figure 16: The 18S rRNA structure of *Plasmodium falciparum* superimposed with the template *T thermophila*

Plasmodium falciparum 18S r RNA energy minimization

Name	18SrRNA.std.egy	18SrRNA.opt.egy
Total Inter Energy	2410217.54112	-34562.06042
Total Intra Energy (-Gamma En)	-16765.12946	-16733.20508
Total Gamma Terms Energy	1433.94236	1473.80569
Total Gap Geometry Penalty	2893.75332	3454.89615
Total Restraint Energy	0.00000	6083.08080
TOTAL STRUCTURE ENERGY	2397780.10735	-46366.56367

Table 4: *Plasmodium falciparum* 18S r RNA Energy optimization Table obtained from results of *RNA 123* that helps minimize the energy from a large positive figure to a more acceptable negative figure that is biologically functional.

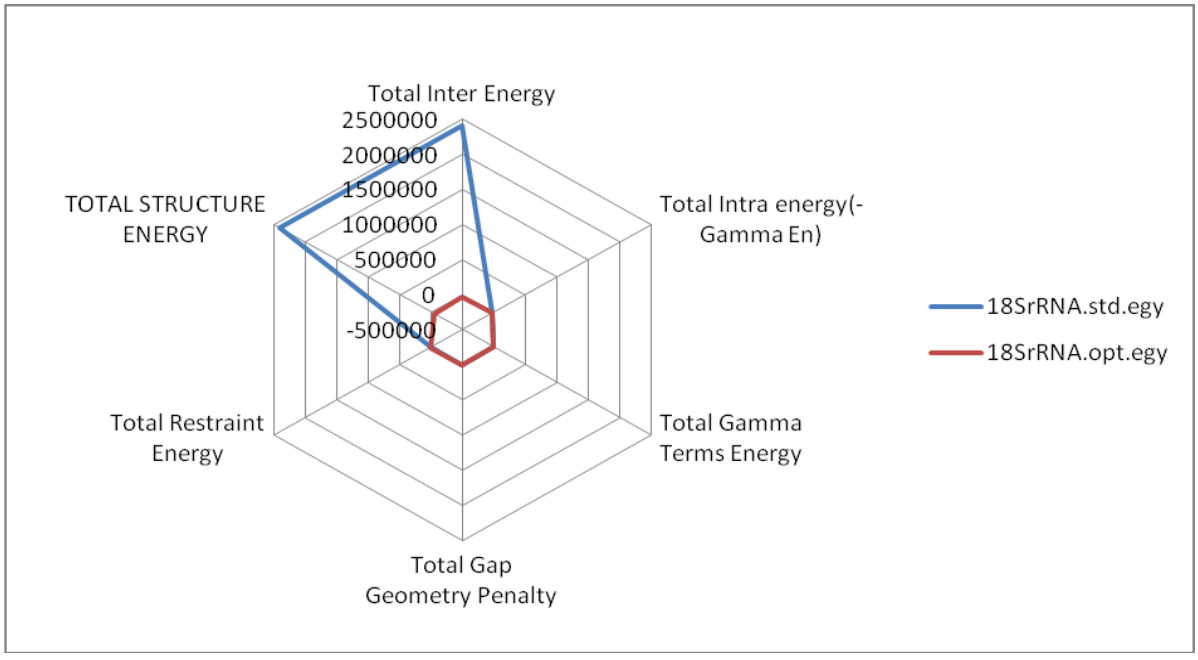


Figure 17: *Plasmodium falciparum* 18S r RNA Energy optimization differences from the positive before energy minimization to a more favorable negative structure.

Structure of Plasmodium Falciparum 40S ribosomal subunit

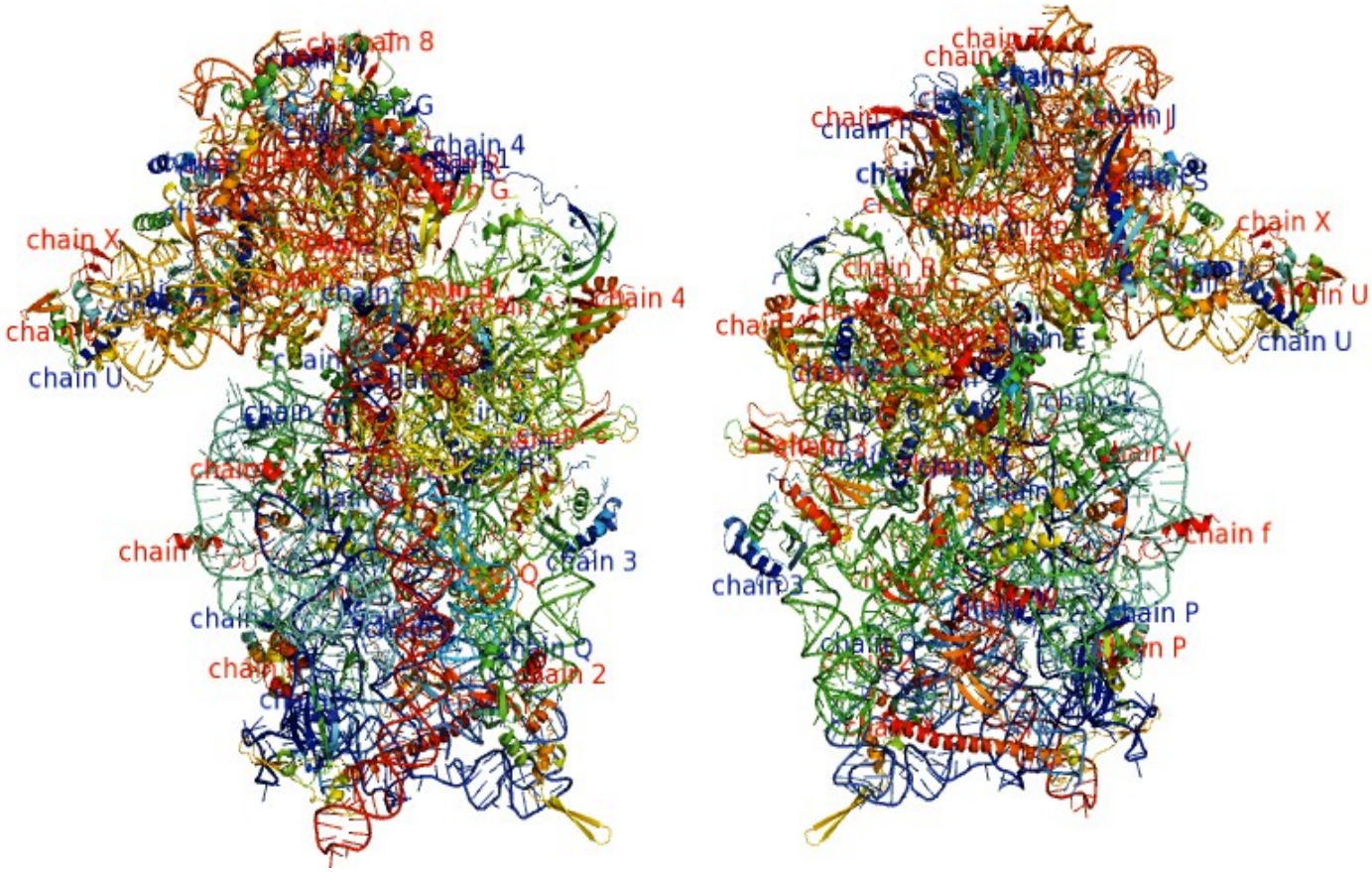


Figure 18 Architectural tertiary structure of plasmodium falciparum 40S front and back view. Shown is the 18S rRNA colored differently depending with domains (5'major –red, Central-green, 3'major-yellow and 3'minor- blue). Also shown are the 40S ribosomal 34 proteins of plasmodium falciparum interacting with the 18S rRNA to make the total subunit.

Chapter Five

5.0 Discussions and conclusion

40S Ribosomal Proteins

The entire process of *Plasmodium falciparum* 40S ribosomal homology modeling was done in a series of stages due to the complexity of the molecule which contains the 18S rRNA and the 34 proteins. Firstly due to the growing repository of eukaryotic ribosome structures solved the process of obtaining a template was conducted and the best homologous molecule was *Tetrahymena thermophila* 40S ribosomal subunit complex with initiation factor 1 solved at 3.9 Å. The ribosome being a much conserved structure due to its functional role in the process of translation contains a more conserved 18S rRNA which may differ between different species on the expansion segments but has a general similar structure across the eukaryotes. Ribosomal proteins are also very highly conserved due to the similar nature of function they play. As shown in Figure 11 all the 34 proteins or plasmodium falciparum were modeled giving a really good structure that could be used to show their relationship with the 18S rRNA. The different tests to ensure that the proteins modeled fall under acceptable parameters which ensure functionality were done and results shown in the validation process shown in Table2 and Table 3 respectively. This give the modeled protein health report which tell one how accurate and reliable ones models are to the same class of proteins in same family. To test how good the proteins are, a Root mean square deviation is performed among other test that add hydrogen's and check for stereochemistry, chirality, side chain among others of each protein. Root mean square deviation is performed in different intervals it basically means

that both structures the model and the template are superimposed therefore different atoms of the both molecules are analyzed to tell how they deviate from each other this is reported as deviations from the template molecule which tells you how the model and template differ from each other. The more the RMSD means that the structures deviate from each other a lot and may be different functionally but a lower RMSD means that the structure are really close to each other and may infers similar functions. RMSD may be performed at different points where the models can be superimposed using PyMOL, Accerlys among others or during validation of proteins the model is superimposed to other models of proteins that in the same family to show homologues and finds the differences in the various methods used in structure determination. Proteins are easier to model compared to the RNA reason being they are smaller in nature and that they have only two dihedral angles compared to the 6 dihedral angles that the RNA.

Proteins are able to perform their functions because of their three dimensional structure they obtain. The structure is stabilized by major weak interactions which include hydrophobic interactions stabilizing soluble proteins globular form; hydrogen bonds and ionic interaction which also help make the structure more stable, covalent bonds in the polypeptide backbone in nature places constrains in the structure.

The three dimensional structure of the polypeptide chain gives the tertiary structure that can be classified into two groups either as fibrous of globular protein. Fibrous group of protein serve structural roles mainly and have simple repeating elements of the secondary structure while the globular proteins tertiary structures are more complicated and contain several types of secondary structures in the same polypeptide chain. Analyzing globular

protein complex structures can be done by examination of stable substructures known as super secondary structures, motifs or folds. Most known proteins have structures generally assembled from combination of this motifs which appear as regions in the polypeptide chains folding stably independently known as domains. Different or similar proteins could come together to form large complexes or protein assemblies.

Since the three dimensional structure gives proteins their functional role as mentioned above this can be destroyed or disrupted if this structure is disturbed but also important to note is that some protein can regain their structure after disruption to take back their natural structure a process dependent only on the amino acid sequence. In the basic unit of life the cell, protein folding involves multiple pathways and for many of them folding is facilitated by the chaperones and chaperonins as the disulphide bonds formation and pro peptide bonds cis –trans isomerization are catalyzed by specific enzymes.

The function of protein involves the interaction with other molecules where the bound molecule is known as a ligand, and site at which this molecule is bound to the protein known as the protein binding site. Conformational changes may happen to a protein when it is bound to the ligand a process which is known as induced fit. This phenomenon could be exploited for the number of proteins that were modeled for *Plasmodium falciparum* 40S subunit to quickly screen available library of possible ligands that would change their conformation and act as intervenes for the parasite. The binding of a ligand to one subunit of a multi subunit protein may affect ligand binding to others subunit. An important thing to note is that the process of binding can be regulated by various factors that alter the binding site of the protein. This protein ligand interaction could change the

organization of the protein either in single or multi subunit protein temporarily in a special degree as shown in some of the protein studied extensively.

Proteins may play another very important part of life which involves specific catalysis of all biochemical reaction and are known as enzymes. All known enzymes are proteins with the exception of catalytic RNAs, requiring co enzymes that are not proteins and co factors for their catalytic functions and are classified according to the reaction they catalyze. Enzymes are highly effective catalyst in enhancing rate of reaction. This enzyme catalyzed reactions are characterized by the formation of an enzyme substrate complex where the substrate binds to the active site of the enzyme. This ensures that the activation energy is reduced and the rate of reaction is enhanced without affecting the equilibrium of the reaction. In cell control of the activities of certain enzymes regulates the activities of metabolic pathways. This is done by feedback inhibition where the end product of the pathway inhibits the first enzyme. Reversible binding of a specific modulator to a regulatory site adjusts the activity of the allosteric enzymes. These modulators may be the other metabolite or substrate itself, and the effect on the modulator may be inhibitory or stimulatory. Comparative interactions among enzymes subunits are a reflection of the kinetic behavior of these allosteric enzymes. Modulation of other regulatory enzymes occurs by covalent modification of a specific functional group that is necessary for the activity, such as the phosphorylation of specific amino groups. Some enzymes such as many proteolytic ones are synthesized as zymogens or inactive precursors, which are later activated by small peptide fragments cleavage. At the important metabolic intersections regulation of enzymes is effected by effectors combinations that allow coordination of activities of the interconnected pathways.

In conclusion these modeled proteins of the *Plasmodium falciparum* could be individually analyzed to further understand their function singly or as combined to show their functional role in the whole 40S ribosomal subunit. This provides an avenue of finding ways that may inhibit its role in translation which would be exploited to give new ways that may help combat the problem that it has caused for many years.

18S ribosomal Ribonucleic acid

The *Plasmodium falciparum* 18S ribosomal ribonucleic acid (rRNA) is a complex structure with 2090 base pairs which form bulk of the 40S subunit of the ribosome accounting for ~60% of its total molecular weight. Due to this complexity modeling was done by segmenting it into the different domains which was much more accurate compared to modeling it as a single unit, then latter combining the domains to form one complex the complete 18S rRNA of *Plasmodium falciparum*. For successful modeling to progress both structures must show similarity for it to be known as comparative modeling. To start with the two secondary structure of both the template (*Tetrahymena thermophila*) and the query (Plasmodium falciparum) show slight differences in the central and the 3'major domains shown in figure 19. This issue was addressed by taking the mismatched regions in the query and fitted them with the matches in the template which appeared as the most logical step to perform. After this was done the template and query were cut at specific regions to ensure similarities at both positions they were cut at. Then from this the segments cut modeling was performed by inserting the both the template coordinates and the query sequence to develop the new model for our query. This was to ensure that in the whole process of modeling the *RNA123* Suite did not have shortfalls that were to bring problems during the process of homology modeling. In figure

12 the 3'monor domain is shown of both the *Plasmodium falciparum* model is coloured in green superimposed to its template coloured brown and blue

In Figure 12 the 3'minor domain was successfully modeled and is shown superimposed to the template to show how homologous they appear but to add to that there are differences which show there are small differences from the template which was expected due to the differences in the primary structure alignments. As observed for both the template and the query 3'minor two dimensional structure show homology and this was done to give a good structure.

In the other 3' major domain despite the differences as stated above with the minor changes made to make homology to incorporate the sequences of the template in the query so to avoid the major changes it was modeled successfully to give the structure in Figure 13. Again shown is a superimposed structure of both template and query. This is shown to be tightly packed structure which plays major role in revelation when finally the structure is combined together.

The central domain follows and is shown in figure 14 again same changes of alteration were performed in the non-homologous regions which made it easier to correctly model the queries sequence correctly. This again shows the structure as a very compact and this is both accounted for by the number of bases in the domain.

Secondary structure diagram of both the template and query 18S rRNA

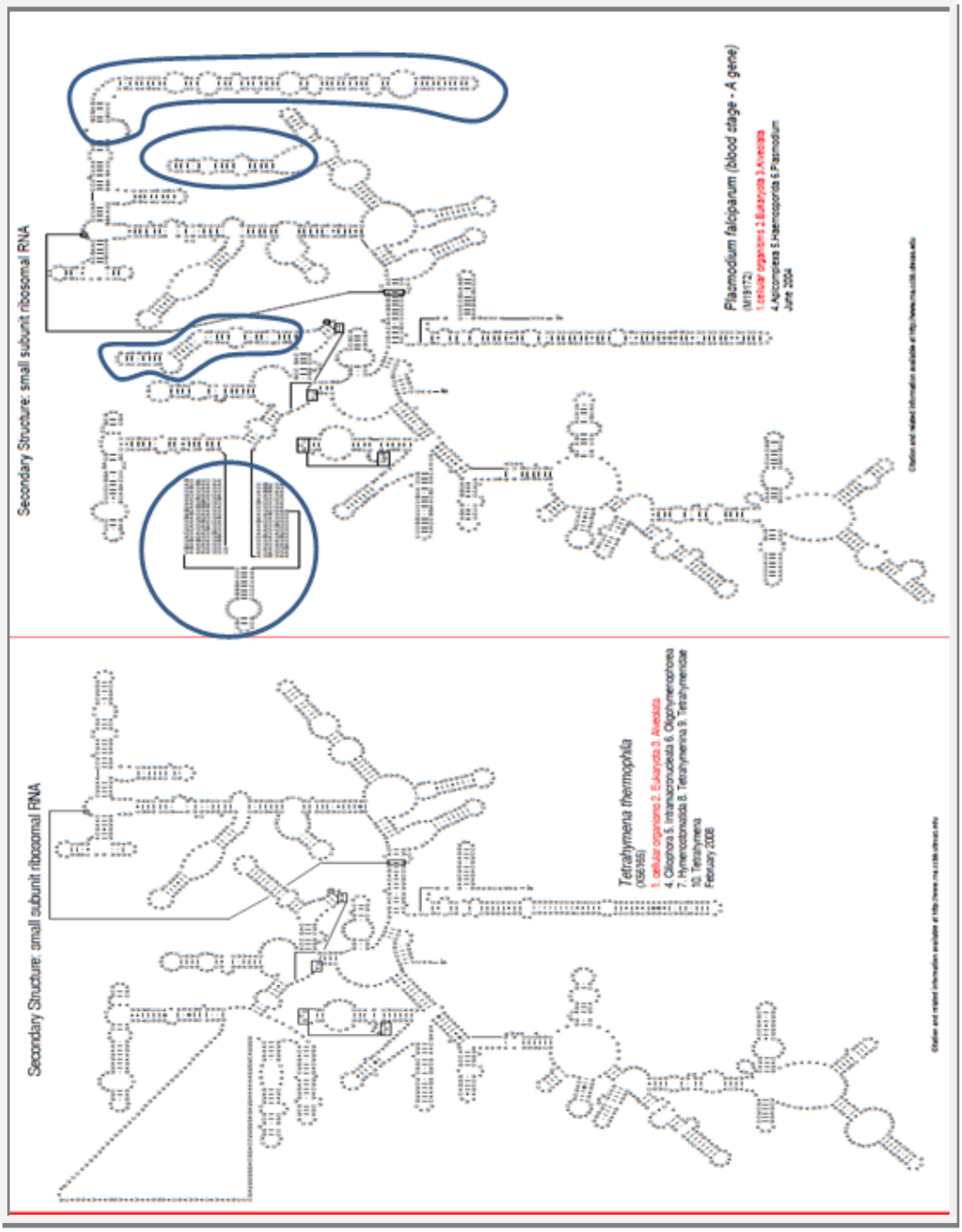


Figure 19 18S rRNA Secondary structure of *Tetrahymena thermophila* and *Plasmodium falciparum* showing the four non-homologous regions between the two structures located at the central and 3'major domain (Cannone, Subramanian et al. 2002).

Finally the fourth domain 5'major was modeled and as shown in Figure 15 above. Again this is image taken using PyMol and the template and models are superimposed together to infer homology. The 5'major domain of the template has 610 base pair compared with the *Plasmodium falciparum.* Although this is observed successful modeling was achieved due to the ability of RNA123 to perform insertions, deletions and cater for the increased number of bases compared to the initial crystal structure. Additional force field calculation is performed to ensure that the structure has minimal energy possible to infer biological function. In conclusion after successfully modeling of the four domains of the complex 18S rRNA of *Plasmodium falciparum* they need to be combined together to form the entire structure of the complete ribosomal ribonucleic acid structure that is to be obtained so as the ribosomal proteins should be added to it to give the 40S subunit. RNA123 suite allows one to join the different segments of the RNA to give a combined structure which can be later optimized to reduce the energy. As shown in figure 16 a combined structure was successfully obtained which was further optimized by running different force fields algorithms that reduce the energy of the structure to bare minimum that conforms to its biological functional morphology. As shown in Table 4 the 18S rRNA standard energies at various positions are compared to the optimized energies of 18S rRNA optimized energy structure. There is a very significant difference between the total inter energies of both structure where the standard total inter energy is a high positive figure whereas ones it's optimized it gives a high negative figure as shown. This tells one that the process of energy minimization at this point is very necessary to ensure biological function of structure. The total intra energy of both structures does not change much and remains a negative figure, same as the Total gamma terms energy. The Total

gap geometry penalty increases slightly with no much significance but the total restraint energy increases from zero to a high of over 6083.08080. the total structure energy of both the standard structure that was obtained after just the combination of the various domains has a higher positive energy that means that its mot biologically viable but after it is optimized and the total structural energy after optimization gives a structure that is highly negative that means that it restraints the prior structure to give a biological allowed structure. All in all the 3 dimensional structure that is obtain after this process was further used to anchor the 34 ribosomal proteins modeled to give the three dimensional structure of the query organism *Plasmodium falciparum.*

The 18S rRNA is a large biological molecule that perform various multiple roles in the coding, decoding, regulation and expresions of genes in cell which is a key for translation in eukaryotes. It is comprised of nucleic acids which appear as single stranded and fold alongside itself to form a highly complex structure that forms basis of the 40S subunit. The rRNA is the component of catalysis in the ribosome and is synthesized in the nucleolus. Recet studies have shown that the rRNA can be used in variuos ways such as futher understanding the role of translation, it being exploited as a drug target for specific organism which have eunique well conserved loops that would be used to dock specific ligands that would stop of cause an effect to the translation process among others. From the structure that was obtained from this work the various loops of the 18S rRNA of *Plasmodium falciparum* would be used to find new intervenes of its propogation and continous cause of a global menace.

Figure 20 shows the combined structure of the 18S rRNA and anchored to it are the 34 proteins. Well illustrated is an architectural tertiary structure of plasmodium falciparum 40S front and back view. Shown is the 18S rRNA colored differently depending with domains (5'major –red, Central-green, 3'major-yellow and 3'minor- blue). Also shown are the 40S ribosomal 34 proteins of plasmodium falciparum interacting with the 18S rRNA to make the total subunit. The various proteins are colored differently to show the various locations on the structure.

The homology structure of the 40S subunit of *Plasmodium falciparum* provides insights into the evolution of the parasite ribosome and into its specific functions. The structure also reveals the interaction of the ribosomal RNA and the proteins providing a step forward towards understanding the eukaryotic parasite translation. Further experiments that include structural, biochemical, genetic and crystallographic studies of the structure described will enable derive required answers of further remaining questions in regards with the assembly, maturation and nuclear export of the subunit and the role it plays in the process of translation. Again the segments that were not modeled due to lack of homology structures could be solved to explain the major differences they would bring to the entire structure and what they might alter to the subunit. This structure could be a very good initial step that helps point a new direction to scientist to quickly understand the process of the translation of the eukaryotic parasites and target the small differences in the process to their host to stop their action and cause a major decline in illnesses they cause.

Chapter Six

6.0 References

Adam B en-Shem, L. J., Gulnara Yusupova, Marat Yusupov (2010). "Crystal Structure of the Eukaryotic Ribosome." Science 330: 1203-1209.

Adriana Verschoor, Suman Srivastava, et al. (1996). "Native 3D Structure of Eukaryotic 80S Ribosome: Morphological Homology with the E. coli 70S Ribosome." Cell Biology 133(3): 495-505.

Albert L. Lehninger, David L. Nelson, et al. (2005). Lehninger-Principles-of-Biochemistry. New York, W. H. Freeman and Company.

Andreeva, A., D. Howorth, et al. (2008). "Data growth and its impact on the SCOP database: new developments." Nucleic Acids Res 36(Database issue): D419-425.

Andrew, E. R. (2009). "Nuclear magnetic resonance." Nuclear Magnetic Resonance. Cambridge Monographs on Physics. ISBN: 9780521114332 1.

Andrew P. Carter, William M. Clemons, et al. (2000). "Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics." NATURE 407 340-348.

Andrew P. Carter, William M. Clemons, et al. (2000). "Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics." Nature 407 340-348.

Armache, J. P., A. M. Anger, et al. (2013). "Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution." Nucleic Acids Res 41(2): 1284-1293.

Armache, J. P., A. Jarasch, et al. (2010). "Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution." Proc Natl Acad Sci U S A 107(46): 19748-19753.

Arnold, K., L. Bordoli, et al. (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics 22(2): 195-201.

Barbara S. Schuwirth, Maria A. Borovinskaya, et al. (2005). "Structures of the Bacterial Ribosome at 3.5 A° Resolution." SCIENCE 310 827-833.

Barty, A., J. Küpper, et al. (2013). "Molecular Imaging Using X-Ray Free-Electron Lasers." Annual review of physical chemistry.

Batey, R. T., R. P. Rambo, et al. (1999). "Tertiary motifs in RNA structure and folding." Angewandte Chemie International Edition 38(16): 2326-2343.

Ben-Shem, A., N. Garreau de Loubresse, et al. (2011). "The structure of the eukaryotic ribosome at 3.0 A resolution." Science 334(6062): 1524-1529.

Beringer, M. (2008). "Modulating the activity of the peptidyl transferase center of the ribosome." RNA 14(5): 795-801.

Bielka (1982). "The Eukaryotic Ribosome." Springer 338(pp 338. Springer-Verlag, Berlin. 1982.).

Bretscher, M. S. (1968). "Translocation in protein synthesis: a hybrid structure model." Nature 218(5142): 675-677.

Brodersen, D. E., W. M. Clemons Jr, et al. (2000). "The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit." Cell 103(7): 1143.

Burwick, N., A. Shimamura, et al. (2011). "Non-Diamond Blackfan anemia disorders of ribosome function: Shwachman Diamond syndrome and 5q- syndrome." Semin Hematol 48(2): 136-143.

Cannone, J. J., S. Subramanian, et al. (2002). "The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs." BMC Bioinformatics 3(1): 2.

Carter, A. P., W. M. Clemons, et al. (2000). "Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics." Nature 407(6802): 340-348.

Chen, V. B., W. B. Arendall, 3rd, et al. (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr 66(Pt 1): 12-21.

Christian MT Spahn, M. G. G.-l., Robert A Grassucci (2004). "Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation." EMBO 23(5): 1008-1019.

Christian MT Spahn, R. B., Narayanan Eswar, Pawel A. Penczek, Andrej Sali, Gunter Blobel, and Joachim Frank (2001). "Structure of the 80S Ribosome from Saccharomyces cerevisiae-tRNA-Ribosome and Subunit-Subunit Interactions." Cell 107: 373-386.

Conte, L. L., S. E. Brenner, et al. (2002). "SCOP database in 2002: refinements accommodate structural genomics." Nucleic acids research 30(1): 264-267.

Crick, J. D. W. a. F. H. C. (1953). "Molecular Structure of Nucleic Acids." NATURE 171: 737-738.

Cundliffe, E. (1987). "On the nature of antibiotic binding sites in ribosomes." Biochimie 69(8): 863-869.

Davis, I. W., A. Leaver-Fay, et al. (2007). "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids." Nucleic Acids Res 35(Web Server issue): W375-383.

Davis, I. W., L. W. Murray, et al. (2004). "MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes." Nucleic Acids Res 32(Web Server issue): W615-619.

Decher, N., M. F. Netter, et al. (2013). "Putative impact of RNA editing on drug discovery." Chemical Biology & Drug Design 81(1): 13-21.

Derek J. Taylor, B. D., Andrew D. Huang, Maya Topf, Eswar Narayanan, Andrej Sali, Stephen C. Harvey, and Joachim Frank (2009). "Comprehensive Molecular Structure of the eukaryotic Ribosome." Science 17: 1591-1104.

Dinman, J. D. (2009). "The eukaryotic ribosome: current status and challenges." J Biol Chem 284(18): 11761-11765.

Ditlev E. Brodersen, W. M. C., Jr.,* Andrew P. Carter,* Robert J. Morgan-Warren,* Brian T. Wimberly,* and V. Ramakrishnan* (2000). "The Structural Basis for the Action of the Antibiotics Tetracycline, Pactamycin, and Hygromycin B on the 30S Ribosomal Subunit." Cell 103: 1143-1154.

Doudna, J. A. and V. L. Rath (2002). "Structure and function of the eukaryotic ribosome: the next frontier." Cell 109(2): 153-156.

Dunkle, J. A. and J. H. Cate (2010). "Ribosome structure and dynamics during translocation and termination." Annu Rev Biophys 39: 227-244.

Frank, J. (2009). "Complexities of the Eukaryotic Ribosome Revealed in New Structural Model." Structure.

Frank, R. K. A. t. a. J. (1999). "Structural studies of the translational apparatus." Current Opinion in Structural Biology 9: 215-221.

Frank Schluenzen, k Ante Tocilj, et al. (2000). "Structure of Functionally Activated Small Ribosomal Subunit at 3.3 A° Resolution." Cell 102: 615-623.

Frederik, P., M. Stuart, et al. (2011). "Perspective and limitations of cryo-electron microscopy." Journal of microscopy 161(2): 253-262.

Fredrick Sijenyi, Pirro Saro, et al. (2012). "The RNA Folding Problems: Different Levels of sRNA Structure Prediction."

Gale, E. F., E. Cundliffe, et al. (1981). The molecular basis of antibiotic action, J. Wiley London.

Garrett, R. (1999). "Mechanics of the ribosome." NATURE 400: 811-813.

Gilly, M., N. R. Benson, et al. (1985). "Affinity labeling the ribosome with eukaryotic-specific antibiotics: (bromoacetyl)trichodermin." Biochemistry 24(21): 5787-5792.

Grabow, W. W., Z. Zhuang, et al. (2013). "The GA-minor submotif as a case study of RNA modularity, prediction, and design." Wiley Interdisciplinary Reviews: RNA.

Grell, L., C. Parkin, et al. (2006). "EZ-Viz, a tool for simplifying molecular viewing in PyMOL." Biochem Mol Biol Educ 34(6): 402-407.

Grigorenko, B., M. Shadrina, et al. (2008). "Mechanism of the chemical step for the guanosine triphosphate (GTP) hydrolysis catalyzed by elongation factor Tu." Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics 1784(12): 1908-1917.

Hansen, J., J. Ippolito, et al. (2004). "synthesis opens new avenues in biotech-nology and medicine, far beyond the development of new antibiotics." Proc. Natl. Acad. Sci. USA 101: 4059-4064.

Hansen, J. L., P. B. Moore, et al. (2003). "Structures of five antibiotics bound at the peptidyl transferase center of the large ribosomal subunit." Journal of Molecular Biology 330(5): 1061-1075.

Hellen, T. V. P. C. U. T. (2006). "Translation interrupted." NATURE STRUCTURAL & MOLECULAR BIOLOGY 13(2): 98-99.

Hendriks, L., A. Goris, et al. (1991). "Phylogenetic analysis of five medically important Candida species as deduced on the basis of small ribosomal subunit RNA sequences." J Gen Microbiol 137(5): 1223-1230.

Hermann, T. (2002). "Rational ligand design for RNA: the role of static structure and conformational flexibility in target recognition." Biochimie 84(9): 869-875.

Hermann, T. and E. Westhof (1998). "RNA as a drug target: chemical, modelling, and evolutionary tools." Current opinion in biotechnology 9(1): 66-73.

Hinnebusch, A. G. and J. R. Lorsch (2012). "The Mechanism of Eukaryotic Translation Initiation: New Insights and Challenges." Cold Spring Harbor Perspectives in Biology.

Ho, D. (2009). About NotePad++, Jan.

Huang, H.-k., H. Yoon, et al. (1997). "GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation inSaccharomyces cerevisiae." Genes & development 11(18): 2396-2413.

J.A. Flohil, G. V., and H.J.C. Berendsen (2002). "Completion and Refinement of 3-DHomology Models with Restricted Molecular Dynamics.Application to target 47, 58 and 111 in the CASP Modeling competition and Posterior Analysis." Proteins 48: 593-604.

Jammie J Cannone, S. S., Murray N Schnare, JamesR Collett, Lisa M D'Souza, Yushi Du, Brian Feng, Nan Lin, Lakshmi V Madabusi, Kirsten M Muller, Nupur Pande, Zhidi Shang, Nan Yu and Robin R Gutell (2002). "The Comparative RNA Web (CRW) Site an online database of comparative sequence and structure information for ribosomal intron, and other RNAs." BMC Bioinformatics 3: 1-31.

Jean-Paul Armachea, Alexander Jarascha, et al. (2010). "Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution." PNAS 107(46): 19748–19753.

Jenner, L., S. Melnikov, et al. (2012). "Crystal structure of the 80S yeast ribosome." Curr Opin Struct Biol 22(6): 759-767.

Jha, S. and A. A. Komar (2011). "Birth, life and death of nascent polypeptide chains." Biotechnology journal 6(6): 623-640.

John Dresios, P. P. a. D. S. (2006). "Eukaryotic ribosomal proteins lacking a eubacterial counterpart Important players in ribosomal function." Molecular Microbiology 59(6): 1651-1663.

Julius Rabl, Marc Leibundgut, et al. (2011). "Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1 " Science 331: 730-736.

Kapp, L. D. and J. R. Lorsch (2004). "The molecular mechanics of eukaryotic translation." Annual review of biochemistry 73(1): 657-704.

Kent L. Redman, a. M. R. (1989). "Identification of the long ubiquitin extention as ribosomal protein S27A." Nature 338: 438-440.

Keqiong Ye, R. J., Jinzhong Lin, Minghua Ju, Jin Peng, Anbi Xu, and Liman Zhang (2009). "Structural organisation of box CD RNA-guided RNA methyltransfrerase." PNAS 106(33): 13808-13813.

Kiefer, F., K. Arnold, et al. (2009). "The SWISS-MODEL Repository and associated resources." Nucleic Acids Res 37(Database issue): D387-392.

Kisselev, L., M. Ehrenberg, et al. (2003). "Termination of translation: interplay of mRNA, rRNAs and release factors?" The EMBO journal 22(2): 175-182.

Klinge, S., F. Voigts-Hoffmann, et al. (2012). "Atomic structures of the eukaryotic ribosome." Trends Biochem Sci 37(5): 189-198.

Krieger, E., G. Koraimann, et al. (2002). "Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field." Proteins 47(3): 393-402.

LaMorte, W. W. (2012). Levels of protein structure. PH709 The Biology of Public Health. L. o. p. structure. Boston, Boston University School of Public Health. 682 × 1000.

Laskowski, R. A., J. A. Rullmannn, et al. (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR 8(4): 477-486.

Leontis, N. B., A. Lescoute, et al. (2006). "The building blocks and motifs of RNA architecture." Current opinion in structural biology 16(3): 279-287.

Levitt, P. K. a. M. (1999). "A brighter future for protein structure prediction." Nature 6(2): 108-110.

Li, G. and X. C. Zhang (2004). "GTP hydrolysis mechanism of Ras-like GTPases." Journal of Molecular Biology 340(5): 921-932.

Lori A. Passmore, T. M. S., David Maag, Drew J. Applefield, Michael G. Acker, Mikkel A. Algire, Jon R. Lorsch, and V. Ramakrishnan (2007). "The Eukaryotic Translation Initiation Factors eLF1 amd eLF1A Induce an open conformation of the 40S Robosome." Molecular Cell 26: 41-50.

Lua, R. C. (2012). "PyKnot: a PyMOL tool for the discovery and analysis of knots in proteins." Bioinformatics 28(15): 2069-2071.

M.Lesk, C. C. a. A. (1986). "The relation between the divergence of sequence and structure in proteins." The EMBO 5(4): 823-826.

Margelevic˘ius, C. V. a. M. (2005). "Comparative Modeling in CASP6 Using ConsensusApproach to Template Selection, Sequence-Structure Alignment, and Structure Assessment." Proteins 7: 99-106.

Maria Selmer, Christine M. Dunham, et al. (2006). "Structure of the 70S Ribosome Complexed with mRNA and tRNA." SCIENCE 313: 1935-1942.

Massa, W. (2011). Crystal structure determination, Springer.

Mathews, D. H. and D. H. Turner (2006). "Prediction of RNA secondary structure by free energy minimization." Curr Opin Struct Biol 16(3): 270-278.

Milne, J. L., M. J. Borgnia, et al. (2013). "Cryo-electron microscopy–a primer for the non-microscopist." FEBS Journal 280(1): 28-45.

MOORE, P. B. and T. A. STEITZ (2003). "After the ribosome structures: How does peptidyl transferase work?" RNA 9(2): 155-159.

Moult, J. (1996). "The current state of the art in protein structure prediction." Current Opinion in structural Biotechnology 7: 422-427.

Murzin, A. G., S. E. Brenner, et al. (1995). "SCOP: a structural classification of proteins database for the investigation of sequences and structures." J Mol Biol 247(4): 536-540.

N. Ban, P. Nissen, et al. (2000). "The Atomic Resolution Crystal Structure of the Large Ribosomal Subunit from Haloarcula marismortui." SCIENCE NSLS Activity Report 2000: 1-6.

Neefs, J. M., Y. Van de Peer, et al. (1991). "Compilation of small ribosomal subunit RNA sequences." Nucleic Acids Res 19 Suppl: 1987-2015.

Nicholas Guex, a. M. C. P. (1997). "SWISS-MODEL and the Swiss -PdbViewer: An environment for comparative protein modeling." Electrophoresis 18: 2714-2723.

Nygard, G. A. a. O. (2004). "Secondary structure of two regions in expansion segments ES3 and ES6 with the potential of forming a tertiary interaction in eukaryotic 40S ribosomal subunits." RNA 10(3): 403-411.

O. Lund, K. F., J. Gorodkin, H. Bohr, J.Bohr, J. Hansen and S. Brunak (1997). "Protein distance constraints predicted by neural networks and probability density functions." Protein Engineering 10(11): 1241-1248.

Ordog, R. (2008). "PyDeT, a PyMOL plug-in for visualizing geometric concepts around proteins." Bioinformation 2(8): 346-347.

Patterson, R. A. a. S. D. (1998). "Current Problems and Technical Solutions in Protein Biochemistry."

Pearson, N. D. and C. D. Prescott (1997). "RNA as a drug target." Chemistry & biology 4(6): 409-414.

Peitsch, M. C. (1996). "ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling." Biochem Soc Trans 24(1): 274-279.

Preethi Chandramouli, M. T., Jean-Franc¸ ois Me´ ne´ tret, Narayanan Eswar, Jamie J. Cannone, Robin R. Gutell, Andrej Sali, and Christopher W. Akey, (2008). "Structure of the Mammalian 80S Ribosome at 8.7 A ° Resolution." Cell: Structure 16 535–548.

Preiss, T. and W. H. M (2003). "Starting the protein synthesis machine: eukaryotic translation initiation." Bioessays 25(12): 1201-1211.

Qi, Y., C. Wu, et al. (2009). "Selection of anti-sulfadimidine specific ScFvs from a hybridoma cell by eukaryotic ribosome display." PLoS One 4(7): e6427.

Ramakrishnan, T. M. S. V. (2009). "What recent ribosome structures have revealed about the mechanism of translation." Nature 461(29): 1234-1242.

Ramakrishnan, V. (2002). "Ribosome Structure and the Mechanism of Translation." Cell 108: 557-572.

Sadee, W., D. Wang, et al. (2011). "Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy." Clinical Pharmacology & Therapeutics 89(3): 355-365.

Sauer, S. D. M. a. R. T. (2007). "The tmRNA System for Translational Surveillance and Ribosome Rescue." Annual Review of Biochemistry 76: 101-124.

Sauer, S. D. M. a. R. T. (2007). "The tmRNA System for Translational Surveillance and Ribosome Rescue." Annu. Rev. Biochem 76: 101-124.

Scheper, G. C., M. S. van der Knaap, et al. (2007). "Translation matters: protein synthesis defects in inherited disease." Nat Rev Genet 8(9): 711-723.

Schwede, T., J. Kopp, et al. (2003). "SWISS-MODEL: an automated protein homology-modeling server." Nucleic acids research 31(13): 3381-3385.

Shi, K., D. R. Houston, et al. (2011). "Crystal Structures of Antibiotic-Bound Complexes of Aminoglycoside 2″-Phosphotransferase IVa Highlight the Diversity in Substrate Binding Modes among Aminoglycoside Kinases." Biochemistry 50(28): 6237-6244.

Simon C. Lovell, M. W., Jane S. Richardson and David C. Richardson (2000). "The Penultimate Rotamer Library." Proteins 40: 389-408.

Spirin, A. S. (1969). "A model of the functioning ribosome: locking and unlocking of the ribosome subparticles." Cold Spring Harb Symp Quant Biol 34: 197-207.

Stephen F. Altschul, T. L. M., Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller and David J. Lipman (1997). "Gapped BLAST and PSI-BLAST a new generation of protein database search programs." Nucleic acids Research 25(17): 3389-3402.

Stephen F. Altschul, W. G., Webb Miller, Eugene W. Meyers and David J. Lipman (1990). "Basic local alignment search tool." Journal of molecular biology 215: 403-410.

Studio, D. (2009). "version 2.5." Accelrys Inc.: San Diego, CA, USA.

Taylor, D. J., B. Devkota, et al. (2009). "Comprehensive molecular structure of the eukaryotic ribosome." Structure 17(12): 1591-1604.

Thierry Lacombe, J. J. G.-G., Jesus de la Cruz, Daniela Roser, Ed Hurt, Patrick Linder and Dieter Kressler (2009). "Linear ubiquitin fusion to Rps31 and its subsequent cleavage are required for the efficient production and functional integrity of 40S ribosomal subunits." Molecular Microbiology 72(1): 69-84.

Thomas Becker, S. B., Alexander Jarasch, Jean- Paul Armache, Soledad Funes, Fabrice Jissinet, James Gumbart, Thorsten Mielke, Otto Berninghausen, Klaus Schulten, Eric Westhof, Reid Gilmore, Elisabet C. Mandon, Roland Beckmann (2009). "Structure of Monomeric Yeast and Mammalian Sec61 Complexes Interacting with the Translating Ribosome." Science 326: 1369-1373.

Thomas, J. R. and P. J. Hergenrother (2008). "Targeting RNA with small molecules." Chemical reviews 108(4): 1171.

Tom L. Blundell, M. S. J., John P. Overington, and Andrej Sali (1987). "Knowledge based modelling of homologous proteins, part I: three-dimensional frameworks derived from the simultaneous superposition of multiple structures." Protein Engineering 1: 377-384.

Varani, G. and W. H. McClain (2000). "The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems." EMBO Rep 1(1): 18-23.

Velagapudi, S. P., S. J. Seedhouse, et al. (2010). "Structure–Activity Relationships through Sequencing (StARTS) Defines Optimal and Suboptimal RNA Motif Targets for Small Molecules." Angewandte Chemie 122(22): 3904-3906.

Vicens, Q. and E. Westhof (2003). "RNA as a drug target: the case of aminoglycosides." ChemBioChem 4(10): 1018-1023.

Vipsita, S., B. K. Shee, et al. (2010). An efficient technique for protein classification using feature extraction by artificial neural networks. India Conference (INDICON), 2010 Annual IEEE, IEEE.

Walter R. P. Scott, P. H. H., Ilario G. Tironi, Alan E. Mark, Salomon R. Billeter, Jens Fennen, Andrew E. Torda, Thomas Huber, Peter Kruger, and Wilfred F. Van Gunsteren (1999). "Biomolecular simulation the gromos96 manual and user guide." Journal of physical chemistry 103: 3596-3607.

Westhof, N. L. a. E. (2012). "RNA 3D Structure analysis and prediction." springer 978(3): 642.

White, C. D. a. S. W. (2000). "Electrons and X-rays gang up on the ribosome." Structure 8(2): 41-45.

William M. Clemons Jr, J. L. C. M., Brian T.Wimberly, John P. McCutcheon, Malcolm S. Capel and V. Ramakrishnan (1999). "Structure of a bacterial 30S ribosomal subunit at 5.5 A resolution." Nature 400: 833-840.

Yuan, J., P. O'Donoghue, et al. (2010). "Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems." FEBS Lett 584(2): 342-349.

Yusupov Marat M, Gulnara Zh. Yusupova, et al. (2001). "Crystal Structure of the Ribosome at 5.5 A Resolution." Science 292(5518): 883-896.

Zhouravleva, G., L. Frolova, et al. (1995). "Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3." The EMBO journal 14(16): 4065.

APPENDICES

7.0 Appendices

7.1 List of computer tools used in this work

This section list the various software tools that were used in each of the levels of the work that was done and goes further to elaborate in detail how they function, and the hard ware that was used to make this work a success.

- **RNA123 Version 2.0.1.3®-** RNA123 as the name suggest is a tool that combines a number of algorithms with the abilities to predict structure, analyze, visualize, structure-based sequence alignment, 3D homology modeling, and de novo modeling of the RNA. Tools for analysis in the RNA 123 suite automatically identify and correct errors in bond length, missing atoms, steric clashes among other errors. Template structures from experimental methods are analyzed automatically identifying hydrogen bonded pairs in it and classify then according to the rules of Leontis-Westhof. The RNA123 suite has a molecular mechanics unique force field that is specifically optimized for RNA allowing it to distinguish native structures from decoys. As shown in Figure 8 RNA 123 has a homology modeling algorithm validated by various sets of RNA targets which include the ribosome, riboswitches and ribozymes. This Validation includes modeling of a known ribosome structure negating all the information generated from its crystal structure, and by using a homologous template a good model was obtained and then compared with the crystal structure from the experimental method. This

resulted to a low energy structure model near to the crystal structure, with correct forecast base paring, stacking and tertiary interactions. To add to this RNA 123 has a sophisticated de novo algorithm for modeling of the RNA 3D structure with fewer than~100nucleotides, which was validated using 25 RNAs and showed an average 0f 3.9 Å RMSD. As per now RNA 123 is available commercially to both academic and industrial licensing runs on windows OS(Fredrick Sijenyi, Pirro Saro et al. 2012).

- **PyMol v0.99®** is a visualization and validation tool which calculates RMSD between atoms if molecules. PyMol is a powerful tool that provides uses ability to view 3D biomolecular images with more than 600settings and 20 representations and can interpreate over 30 different file formats form basic pdb files, multi sdf files to volumetric electron density maps. Its easy to use due to its graphical user interface allowing any type of user to create stunning 3d images from favorite file formats. Saving Images and movies can be done in a cross-platform session file, and viewed exactly as intended, with assurance that object position, colour of atom, representation of molecule,state of molecule, frame and movie viewingis precise. The 20 different ways that data ca be represented include Spheres which provides a CPK-like view,Mesh and surface which provide views that are more volumetric, Sticks and lines which put bond connectivity emphasis, cartoon and ribbon which offer popular ans easy representation for the identification of the secondary structure topology (Grell, Parkin et al. 2006; Ordog 2008; Lua 2012).
- **Discovery studio Accelrys 2.50®** is a suite for small molecule and macromolecule simulation software developed and distributed by the Accelrys

Company specializing in scientific software production. Discovery studio is a client server software suite, that is built on a pipeline visual programming, entailing computational chemistry, computational biology, chemiinformatics, molecular simulation and quantum mechanics. Discovery studio major uses involves the development of novel therapeutics medicines which include small molecule drug, therapeutic antibodies, vaccines, synthetic enzymes and others such as consumer products. This property makes it be used over a number of academic and commercial entities such as pharmaceutical, biotech and consumer goods industries. It can run in both windows Red Hat and suse Linux client servers. Discovery studio has a strong academic collaborative program, supportive scientific research to its suite, making a number of software algorithms originally developed in the scientific community which includes CHARMM, MODELLER, DELPHI, ZDOCK, and DMol3 among others (Studio 2009).

- **MolProbity**- Is a server that offers general purpose web service on quality validation of 3D structures of proteins, nucleic acids and other complex molecules. It was developed by the Richardson group and provides all atom detailed contact analysis of any steric problem that may be found within the molecule that is being evaluated, it calculates and displays the H-bond and van der walls contacts in the interfaces between various components. An important step in this process of validation is the ability to add and fully optimize all hydrogen atoms, both polar and non-polar. MolProbity is a very valuable structure validation tool which after the above steps relays results in multiple forms as overall numeric scores, lists, downloadable PDB and graphical files. The results

are as informative, easily manipulated 3D kinemage graphics shown online in kiNG tool viewer. This server is available for all and is available freely to all users at http://kinemage.biochem.duke.edu (Davis, Murray et al. 2004; Davis, Leaver-Fay et al. 2007; Chen, Arendall et al. 2010).

- **PROCHECK version 3.6.2 ®** Is a tool developed by European Molecular Biology Laboratory that provides an idea of the stereochemistry quality of the protein chains in a given input PDB structure. It analyses the structure highlighting the protein regions appearing to have unusual geometry and give an overall detailed analysis of the structure as a whole. PROCHECK is easy to use where it allows one to input the protein PDB file and then one can navigate through to find analysis required for validating protein. It provides four major analytical ways of providing results either as Plot description under which falls; main ramachandran plot, all residue ramachandran plots, All residue chi1-chi2 plots, main chain parameters, side chain parameters, residue properties plot, main chain bond lengths, Main chain bond angles, RMS distances from planarity, and distorted geometry. Print out; residue by residue listing under 3D view; Torsion angle G factor in 3D and finally Miscellaneous; G factor. Result can be evaluated to tell one of finding that is provided (Laskowski, Rullmannn et al. 1996).

- **Swiss-Model server®-** Is a bioinformatics structural tool that is set up as a web server that is dedicated to 3D protein homology modeling. Its fully automated and freely accesible via the ExPASy web server, or from swiss Pdb viewer Deepview. Swiss Model comprises three components integrated tightly. This are SWISS

MODEL pipeline a combination of software tool and databases that allow for automated modeling of protein structures. Next is SWISS MODEL workspace a graphical web based user workbench and finally SWISS MODEL Repository which is a database that is continously updated of homology models for set of model organism proteomes of high biomedical intreast (Peitsch 1996; Nicholas Guex 1997; Arnold, Bordoli et al. 2006; Kiefer, Arnold et al. 2009).

- **SCOP** means structure classification of proteins and is used to classify proteins as the name sugests into five levels; class, fold, super family, family. Division of classes is done as alpha, all beta, alpha or beta alpha and beta domains. Super family for example does classify by convergent evolutin where there is unrelated function but same structure as the evolutionaly disctint lineage. Proteins that are orthologous with same functions but different organism; related descent from duplicate common ancestors but different function or paralogous proteins . Family comes finally where diferentiation of these paralogous proteins is done
- **YASARA®** is an acronym of Yet Another Scientific Artificial Reality Application which is a molecular graphics, simulation and modeling software available for windows, linux and Mac OS X users. It has a simple graphical user interface thatallows users to focus on their goal . Its powered by a new developmental frame work providing perfomance better than traditional software the portable vector language. This allows one to visualize much larger proteins and interactive real time simulations with highly accurate forcefields on a standard pc. Is based on academic world hence all methods are reviewed by journals (Krieger, Koraimann et al. 2002).

- **Notepad++** Is a source code editor that supports several languages, runs in a windows platform governed by GPL licence use and available for free. Its based on scintilla a powerful editing component and is written in c++ using pure Win32 API and STL ensuring higher execution speed and smaller program size. Notepad++ is user friendly although optimised in many routenes (Ho 2009).

7.2 Sequences used as template

Crystal structure of the Eukaryotic 40S Ribosomal subunit of *Tetrahymena thermophila* used in this work as Fasta format.

>2XZM:1|PDBID|CHAIN|SEQUENCE

MNTEETTKARIMDVLGKTGSRGGITQVRVVLISGKEEGRQLIRNVKGACRVGDVLELMECEREAR
RLR

>2XZM:2|PDBID|CHAIN|SEQUENCE

MGISRDSKHKRRATGGRMPVHRKKRAFEKGRPISMTKLTTQSTTITEKRRIRPVRVRGGHLKFRAL
RLCEGNFSWGSENITRKTKILDVKYNATNNELVRTKTLVKNSIVEIDSTPFREWYKLHYGIDLGLKK
DRTVLGNKEKSRHVQKRVKRTKAQALEKNIEEQFVSQRILACITSRPGQSGRADGYILEGKELEFYI
RKLQSKKK

>2XZM:3|PDBID|CHAIN|SEQUENCE

MALAKFHKKQSLTKIEEQVGSALVELQNTHPDLKTSLESIILTQVKEFQINKTNKKAKSAVLIYVHF
QSYRVLLSAARKLIIELEKKLKQIVFFTAQRKIESKWVKEHKSQQRPRSRCLTYVYDALLDDLLLPS
TLIGKRIRARLDGTSFYRIQLDQNDRDFLEEKLDAITHIYKTVTTREVTFEFKEDKTFHTFKK

>2XZM:4|PDBID|CHAIN|SEQUENCE

MTQGKNPGLKAGGGKKGAKKRTIDPLSRKEWYDFRAPIPFSSKSFGKTLVTKSSGNRIASEEIKGR
VVESTLADLKDNSNDKAWRKVKLVIDEVDGRNAKTSFYGLDITRDRLCSMIRKWQTLIEARVDCK
TNDGYIIRVFTLAFTKKTSAGKQSSTSTCYAKSSQVRAIRRKINTFITNEAAKLGIAEFSKNLIGEDYT
KKIEKETKNIFPLQNITIRKVKVLKRPKLDATKIAELYSHEKKGEKATGRDGAPEEQAAQNLLAQ

>2XZM:5|PDBID|CHAIN|SEQUENCE

MPVKRRNAGRSQKNRGHTRTVPCTNCGRQVAKDKAVKRYTVRDMVDPSSKRDIQQKLAFENEK
QGIPKLYVKLQYCISCAIHSRVVRVRCAEDRRIRRPPIRNRKPQVVKTDATAPKKQ

>2XZM:6|PDBID|CHAIN|SEQUENCE

MERDLLNPIYEEEKQKNKFKRLIQAPNSYFMDVKCAQCQNIQMIFSNAQSTIICEKCSAILCKPTGG
KVQIQAGCAFKIKN

>2XZM:7|PDBID|CHAIN|SEQUENCE

MVHVLKATKIRIYKQLLQDGVFVLKKDFEGHHEETGVPNLHCYILVRSLKDRGFLEEIFNWGFTYY
YLNKEGCEYLKTKLGISADNVIPKTFKASNVNFISKEEDEEERPRRQFNKGGRTGERDGRNKRGVG
RGTRREGEEAAKEEGAAETAQGNQETPAQE

>2XZM:8|PDBID|CHAIN|SEQUENCE

MGKQPAAGQKKTKEAIAKAAQAKKGGKKKWTKGKAKDKVNHAVFIEKKNVESIINNPSKVGKV
LTVSTVVEKLKVNGSLARQLMRTMADRKLVEKVAKNGNQWVYSVIGGVKEDKTAAPAAAGGKK
QQQSKKGAAADKEEVQA

>2XZM:9|PDBID|CHAIN|SEQUENCE

MQVQVKTLEGETKIYTLEQGTSVLDLKSQISQDMGFEIDMMTLVNNGFIAPNTELVTDDVTYYLSL
KLLGGKKKKKKKSYTTKKKTKHRHVHTKLGALAFYKLENNGKVSLQQKGCPKCGPGIFMAKHY
DRHYCGKCHLTLKIDXXXXXXXXXXXXXXXXXXXXXXXXXDAGAGGAGKKDAKKGKKGKK

>2XZM:A|PDBID|CHAIN|SEQUENCE

AACCUGGUUGAUCCUGCCAGUUACAUAUGCUUGUCUUAAAUAUUAACCCAUGCAUGUGCC
AGUUCAGUAUUGAACAGCGAAACUGCGAAUGGCUCAUUAAAACAGUUAUAGUUUAUUUGA
UAAUUAAAGAUUACAUGGAUAACCGAGCUAAUUGUUGGGCUAAUACAUGCUUAAAAUUCC
GUGUCCUGCGACCGGAACGUAUUUAUUAGAUAUUAGACCAAUCGCAGCAAUGUGAUUGAG
AUGAAUCAAAGUAACUGAUCGGAUCGAGGUUUACCUCGAUAAAUCAUCUAAGUUUCUGCC
CUAUCAGCUCUCGAUGGUAGUGUAUUGGACUACCAUGGCAGUCACGGGUAACGGAGAAUU
AGGGUUCGAUUCCGGAGAAGGAGCCUGAGAAACGGCUACUACAACUACGGUUCGGCAGCA
GGGAAGAAAAUUGGCCAAUCCUAAUUCAGGGAGCCAGUGACAAGAAAUAGCAAGCUGGGA
AACUUACGUUUCUACGGCAUUGAAAUGAGAACAGUGUAAAUCUCUUAGCGAGGAACAAUU
GGAGGGCAAGUCAUGGUGCCAGCAGCCGCGGUAAUUCCAGCUCCAAUAGCGUAUAUUAAA
GUUGUUGCAGUUAAAAAGCUCGUAGUUGAACUUCUGUUCAGGUUCAUUUCGAUUCGUCGU
GUGAAACUGGACAUACGUUUGCAAACUAAAAUCGGCCUUCACUGGUUCGACUUAGGGAGU
AAACAUUUUACUGUGAAAAAAUUAGAGUGUUCCAGGCAGGUUUUAGCCCGAAUACAUUAG
CAUGGAAUAAUGGAAUAGGACUAAGUCCAUUUUAUUGGUUCUUGGAUUUGGUAAUGAUUA
AUAGGGACAGUUGGGGGCAUUAGUAUUUAAUAGUCAGAGGUGAAAUUCUUGGAUUUAUUA
AGGACUAACUAAUGCGAAAGCAUUUGCCAAAGAUGUUUUCAUUAAUCAAGAACGAAAGUU
AGGGGAUCAAAGACGAUCAGAUACCGUCGUAGUCUUAACUAUAAACUAUACCGACUCGGG
AUCGGCUGGAAUAAAUGUCCAGUCGGCACCGUAUGAGAAAUCAAAGUCUUUGGGUUCUGG
GGGAAGUAUGGUACGCAAGUCUGAAACUUAAAGGAAUUGACGGAACAGCACACCAGAAGU
GGAACCUGCGGCUUAAUUUGACUCAACACGGGGAAACUCACGAGCGCAAGACAGAGAAGG

GAUUGACAGAUUGAGAGCUCUUUCUUGAUUCUUUGGGUGGUGGUGCAUGGCCGUUCUUAG
UUGGUGGAGUGAUUUGUCUGGUUAAUUCCGUUAACGAACGAGACCUUAACCUGCUAACUA
GUCUGCUUGUAAAUAACAGGUUGUACUUCUUAGAGGGACUAUUGUGCAAUAAGCCAAUGG
AAGUUUAAGGCAAUAACAGGUCUGUGAUGCCCCUAGACGUGCUCGGCCGCACGCGCGUUAC
AAUGACUGGCGCAAAAAGUAUUUCCUGUCCUGGGAAGGUACGGGUAAUCUUAUUAAUACC
AGUCGUGUUAGGGAUAGUUCUUUGGAAUUGUGGAUCUUGAACGAGGAAUUUCUAGUAAGU
GCAAGUCAUCAGCUUGCGUUGAUUAUGUCCCUGCCGUUUGUACACACCGCCCGUCGCUUGU
AGUAACGAAUGGUCUGGUGAACCUUCUGGACUGCGACAGCAAUGUUGCGGAAAAAUAAGU
AAACCCUACCAUUUGGAACAACAAGAAGUCGUAACAAGGUAUCUGUAGGUGAACCUGCAG
AUGGAUCAUUA

>2XZM:B|PDBID|CHAIN|SEQUENCE

MATQRKQDDIKRLLASNCHQATINLNNQMKRYISHKGVNGIHYINIEETWQKIKLAARVIAAVQHP
EDVMVVCSRIYGQRAAIKFAGYTHCKSTSSSRWTPGTLTNYQTLKYEEPRVLIVTDPRSDFQAIKEA
SYVNIPVIALCDSDSPLAYVDVVIPCNNRSTESISMIYWMIAREVKILRGELSKDEEWEVMVDLFYH
KTLPTAEQKEAEEEEGAEGAEEKVAEVKEGEAEQNTDNKNW

>2XZM:C|PDBID|CHAIN|SEQUENCE

MDKTTRAINKKKKFVADGVFNAELHSFFSKSLQDAGYAGIEVRRTPTKTEIRIKATKPQQVIGVEGK
KHKELTQFLQKRFGYSDDQIQIWAEPIKFKGLCASAQVEAMNYKLLKDVPVRLAANYIIKSVIQDG
AKGCEIIISGKLKQQRAKTMKFKQGYMICTGQPKNDYIDVAVRHVFFKQGIMGVKVKIMLPYEPNP
AKKFGVKTPIPDNVIIHPPKQITDDKEIRTAVEQQQPAQPEQQQ

>2XZM:D|PDBID|CHAIN|SEQUENCE

MGKTYINTSKTYSTPRRPYEKERLDNEMKLIGTFGLKNKREVWRVQMILAKFRKAARELLTLDPK
DPRRLFEGSALLRRMFKYGLLSEQERKLDYVLGLTIHKLMERRLQTRVFKLNLANSIHHSRVLIRQ
RHIKVGKNLVNVPSFMVRTDSEKSIDFASTSPLGGGREGRTKRKNAKKSA

>2XZM:E|PDBID|CHAIN|SEQUENCE

MAQRTEKRGFGNKKGPRGDRRGDKNDWQPQTKLGRLVKYGKISSLDEIFKYSIPIKEPEIIDHFYPK
NKDVQAEHKLMEEVLQITPVQKQTQAGQRTRFKGFVVVGDSNGHIGLGWKVAKEVQGAIKGAIT
HAKLNMVPVRKGYWGNKIANAHTIPQKITGKSGSVRIRLVPAPRGTGIVAAPIPKKVLQFAGVQDI
YTSSQGCTRTRGNFLKATYYALANTYRYLTPDFWGKPEDNELPFETFSEFLHTKAQKKEPKQYENK
REKKHFDRPRREHGDRQPRGEKKTEEVPATEAQ

>2XZM:F|PDBID|CHAIN|SEQUENCE

MNDNDFQFENNIDDFQTHIHIRVEQRRGRKCFTTVEGIPPEFDYEKIMKYWKKWLSCNATIVEEDE
GKKVIKLNGDHRNQIQQFLSEEGIAAVDNITIHGI

>2XZM:G|PDBID|CHAIN|SEQUENCE

MSVEKGFATQPKLFGKWNYDEVKIQDPCFQNYIACTTTKSQVFVPHTAGRYQVKKFRKTQCPIVE
RLIGTLMFHGRNAGKKALCIKVVKNAFEIIHLVTGRNPLEVFVGAVQNAGPREDSTRIGTAGVVRK
QAVDVAPMRRVNLAIYFIIKGCRESAFKSMRSIAETLADEIINAEKNNTQSSWAIRKKDEIEKVAKGN
R

>2XZM:H|PDBID|CHAIN|SEQUENCE

MVKVNILAECLKDLVNAEKMGKKQVLLRPVSKVVLKFLRIMQKNGYIGEFEVIDDHRSKKVVVE
LIGRINKCGVISPRYDVPLADFEKWTNNILPSRQFGNVVLTTTYGILTHEECRKRHTGGKILGFFY

>2XZM:I|PDBID|CHAIN|SEQUENCE

MSQQKPQLVQTFGRKKNAVAVASVRPGKGLLKVNGSPIDMINPQILQAKIYEPILLLGQQKFANLDI
RIRVRGSGYTSQVYAIRQALSKGIVAYHAKYVDENSKREIKEQLMQYDRSLLVADPRRMEPKKCGG
RGARSKMQKAYR

>2XZM:J|PDBID|CHAIN|SEQUENCE

MDATKETKKGEEEVKKVRVRITLTCKNLKSVEKATSEIVSRAKKTEQVEVKGPVRMPTKTLVITVR
KSPCGEGSKTWDRFEMRIYKRIIDLTCNVPDVKTITNFRIDPGVEIELTMTADQ

>2XZM:K|PDBID|CHAIN|SEQUENCE

MAPKKTAPVQKEVISYGPPNVGANENVFGVCHIMATWNDTFIHVTDLSGRETLVRVTGGMKVKA
DREESSPYAAMQAAIDVVNRCKELKINALHIKLRAKGGVETKQPGPGAQSALRALARSGMKIGRIE
DVTPIPTDSTRREGGRRGRRL

>2XZM:L|PDBID|CHAIN|SEQUENCE

MGVGKPRGIRAGRKLARHRKDQRWADNDFNKRLLGSRWRNPFMGASHAKGLVTEKIGIESKQPN
SAVRKCVRVLLRKNSKKIAAFVPMDGCLNFLAENDEVLVAGLGRQGHAVGDIPGVRFKVVCVKGI
SLLALFKGKKEKR

>2XZM:M|PDBID|CHAIN|SEQUENCE

MSFVIEKESDFKYIHRILNTNIDGKRITPIALTGIRGIGRRFAYIICKVLKIDPNARAGLLTEDQCNKIT
DLIADPEAHGIPTWLLNRINDFKDGKNYQMASNTLDTKMREDLERLKKIKSHRGLRHFWGLKVR
GQHTKTSGRHGVVCGVVRKNK

>2XZM:N|PDBID|CHAIN|SEQUENCE

MPNKLWRTHPRNYGKDSKECRVCGARQGLITKYEMMTCRRCFREQAPHIGFVKYR

>2XZM:O|PDBID|CHAIN|SEQUENCE

MGRMQMKGKGKGISGSALPFKRRSPKWLHMTPSTVVDLSVKLAKKGLTPSQIGVILRDQHGIPQV
RFLTGQKILRILKKNGCAPQLPEDLYFLIKKALSIRKHLEKNRKDKDSKYRLILVESRIHRLSRYYKL
NQKLPPKWKYNAQTASALVQ

>2XZM:P|PDBID|CHAIN|SEQUENCE

MTIVIRTKKILVNPLLSRRQLSLDVLHPDSPTASKEKIREELAKQLKVDARNVVVYGFSTQYGGGKS
TGFALVYDNQQYLLKYEPNYRLRKVKILGEKPNTRRSFKELKRKIKRTSGKAITKLLSEKKGDTWA
SVQSKKSDHLKNFVAK

>2XZM:Q|PDBID|CHAIN|SEQUENCE

MDTQIQRAYQKQDGVFLNSKKLLAKKTSAGVRYYKNIGLGFKTPKEAIEGTYVDKKCPFTSNLSIR
GKIIKGLVISTKMNRTVIIRRDYLHYVRKYNRYEKRHRNIPVHISPCFSVKEGDILVAGQCRPISKTV
RFNALQVVPNEIIGSVRKQFLLF

>2XZM:R|PDBID|CHAIN|SEQUENCE

MADNSSLDIQVVKRGILEGHSDWVTSIVAGFSQKENEDSPVLISGSRDKTVMIWKLYEEEQNGYFG
IPHKALTGHNHFVSDLALSQENCFAISSSWDKTLRLWDLRTGTTYKRFVGHQSEVYSVAFSPDNRQI
LSAGAEREIKLWNILGECKFSSAEKENHSDWVSCVRYSPIMKSANKVQPFAPYFASVGWDGRLKV
WNTNFQIRYTFKAHESNVNHLSISPNGKYIATGGKDKKLLIWDILNLTYPQREFDAGSTINQIAFNP
KLQWVAVGTDQGVKIFNLMTQSKAPVCTIEAEPITKAEGQKGKNPQCTSLAWNALGKKLFAGFTD
GVIRTFSFETSAN

>2XZM:S|PDBID|CHAIN|SEQUENCE

MADNKKSIKKFTFRGKGLEELTALASGSNSEKLISDELAALFDAKTRRRVKRGISEKYAKFVNKVR
RSKEKCPAGEKPVPVKTHYRSMIVIPELVGGIVGVYNGKEFVNVEVKFDMIGKYLAEFAMTYKPTT
HGKTGAAGGKGK

>2XZM:T|PDBID|CHAIN|SEQUENCE

MSNNQSFFTVKDVAAADFIREYASHLKKANKLSIPEFTQWTTTSVARELAPQDSDWVYIRTAALAR
KVYLKPHTGISTLKHIFGSNKDRGNLRNKHQACHGKILRWALKSLEDLKIIRKDKNSATKKFSRVIT
KEGMTELNRIATQIAIKQRQAK

>2XZM:U|PDBID|CHAIN|SEQUENCE

MADQNQQLNEVLAKVIKSSNCQDAISKGLHEVLRTIEAKQALFVCVAEDCDQGNYVKLVKALCA
KNEIKYVSVPKRASLGEYLGHFTANAKGEIKKVKGCSSLAIRKYAPEITEDEKKIIEGALKA

>2XZM:V|PDBID|CHAIN|SEQUENCE

MGRVRTKTVKRAAKSLIEHYYSKLTNDFHFNKKILSEVAQVPSKRLRNKIAGFATHLMKRIQKGPV
RGISLKVQEEERERRLDYVPEKSIIDIEKVTIDNETKEMLKKLGYQNIPGLVSASAPANKKRGQ

>2XZM:W|PDBID|CHAIN|SEQUENCE

MARGPKKHLKRINAPKSWMLNKLGGIWATRPSQGPHKLRESLPLSVLLKERLNYALNGRDVTLIL
NDKEQNVFVDGKVRRDKGYPTGLMDVVRIEKTDQSFRILYDTKGRFVLKSLSKEEAKYKLLKVTA
KAIGPNQIPYIVTHDSRTIRFPNPEIKIGDTLKYDLVNNKIENFAHLESGNVCYIQQGNNIGRVGIIQHI
EKHQGSFDICHVKDAKGNAFATRLGNIFVLGQGKKSWIELPSGDGVRETILEERKRKFSY

>2XZM:X|PDBID|CHAIN|SEQUENCE

MGRMHGTLAKAGKVRKQTPKVEKKDKPRKTPKGRSYKRILYNRRYAPHILATDPKKRKSPNWHA
GKKEKMDAAANPVKKD

>2XZM:Y|PDBID|CHAIN|SEQUENCE

MKFNISYPLTGAQKCIEIDDDKKCNIFMDKKMGQEVEGDTLGDEFKGYVFKIAGGNDKDGFPMK
QGVMVRGRVRLLLSEGHSCFTSRRSGFRKRKSVRGCIVGPDMRVLALQIVKKGVAEIDGLTTVTLP
RKLGPKRANNIKKLFGLKKEDDPILIKKSVIRRTFKTAKGKDRTKCPKIQRLITPERILRKKVYKAEK
TQRYVKTNAAKEEYEKFLSEWKKQRAAKAHAASAPVVEAPKKVEAPKKVDPKAAKTTPAATKAT
PAATKVAPKTQAAKTTTPAPAVKDAKKTTKK

>2XZM:Z|PDBID|CHAIN|SEQUENCE

MNSGRANQRSKPMTGLINDKKEKIDAYLPRKCDWSNKLIFSNDQSSVQIAIAEVGENGQATGSKTN
VVLCGSVRSKGEAHIALENILRERGLYPIQE

7.3 Sequences of the 40S Plasmodium falciparum subunit used in this work

>gi|160642|gb|M19172.1|PFARGEA P.falciparum 18S ribosomal RNA in asexual parasites

AACCTGGTTGATCTTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCAAGTGAAAGTA
TATATATATTTTATATGTAGAAACTGCGAACGGCTCATTAAAACAGTTATAGTCTACTTGACATTTT
TATTATAAGGATAACTACGGAAAAGCTGTAGCTAATACTTGCTTTATTATCCTTGATTTTTATCTTT
GGATAAGTATTTGTTAGGCCTTATAAGAAAAAAGTTATTAACTTAAGGAATTATAACAAAGAAGT
AACACGTAATAAATTTATTTTATTTAGTGTGTATCAATCGAGTTTCTGACCTATCAGCTTTTGATG
TTAGGGTATTGGCCTAACATGGCTATGACGGGTAACGGGGAATTAGAGTTCGATTCCGGAGAGG
GAGCCTGAGAAATAGCTACCACATCTAAGGAAGGCAGCAGGCGCGTAAATTACCCAATTCTAA
AGAAGAGAGGTAGTGACAAGAAATAACAATGCAAGGCCAATTTTTGGTTTTGTAATTGGAATG
GTGGGAATTTAAAACCTTCCCAGAGTAACAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGG
TAATTCCAGCTCCAATAGCGTATATTAAAATTGTTGCAGTTAAAACGCTCGTAGTTGAATTTCAA
AGAATCGATATTTTATTGTAACTATTCTAGGGGAACTATTTTAGCTTTTGGCTTTAATACGCTTCCT
CTATTATTATGTTCTTTAAATAACAAAGATTCTTTTTAAAATCCCCACTTTTGCTTTTGCTTTTTTG
GGGATTTTGTTACTTTGAGTAAATTAGAGTGTTCAAAGCAAACAGTTAAAGCATTTACTGTGTTT
GAATACTATAGCATGGAATAACAAAATTGAACAAGCTAAAATTTTTTGTTCTTTTTTCTTATTTTG
GCTTAGTTACGATTAATAGGAGTAGCTTGGGGACATTCGTATTCAGATGTCAGAGGTGAAATTCT
TAGATTTTCTGGAGACGAACAACTGCGAAAGCATTTGTCTAAAATACTTCCATTAATCAAGAAC
GAAAGTTAAGGGAGTGAAGACGATCAGATACCGTCGTAATCTTAACCATAAACTATGCCGACTA
GGTGTTGGATGAAAGTGTTAAAAATAAAAGTCATCTTTCGAGGTGACTTTTAGATTGCTTCCTTC
AGTACCTTATGAGAAATCAAAGTCTTTGGGTTCTGGGGCGAGTATTCGCGCAAGCGAGAAAGTT
AAAAGAATTGACGGAAGGGCACCACCAGGCGTGGAGCTTGCGGCTTAATTTGACTCAACACGG
GGAAACTCACTAGTTTAAGACAAGAGTAGGATTGACAGATTAATAGCTCTTTCTTGATTTCTTGG
ATGGTGATGCATGGCCGTTTTTAGTTCGTGAATATGATTTGTCTGGTTAATTCCGATAACGAACG
AGATCTTAACCTGCTAATTAGCGGCGAGTACACTATATTCTTATTTGAAATTGAACATAGGTAACT
ATACATTTATTCAGTAATCAAATTAGGATATTTTTATTAAAATATCCTTTTCCCTGTTCTACTAATAA
ATTGTTTTTTACTCTATTTCTCTCTTCTTTTAAGAATGTACTTGCTTGATTGAAAAGCTTCTTAGA
GGAACATTGTGTGTCTAACACAAGGAAGTTTAAGGCAACAACAGGTCTGTGATGTCCTTAGAT
GAACTAGGCTGCACGCGTGCTACACTGATATATATAACGAGTTTTTAAAAATATGCTTATATTTGT
ATCTTTGATGCTTATATTTTGCATACTTTTCCTCCGCCGAAAGGCGTAGGTAATCTTTATCAATATA
TATCGTGATGGGGATAGATTATTGCAATTATTAATCTTGAACGAGGAATGCCTAGTAAGCATGATT
CATCAGATTGTGCTGACTACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAA
AGATATGATGAATTGTTTGGACAAGAAAAATTGAATTATATTCTTTTTTTTCTGGAAAAACCGTA
AATCCTATCTTTTAAAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGG
ATCATTA

>gi|124809606|ref|XP_001348622.1| 40S ribosomal protein S2, putative [Plasmodium falciparum 3D7]

MEDRGGFSRGFGRGVRGTRGRGGRGARGRGRGSAEDDLKNWVPVTKLGRLVKEGKIVSIEEIYL
HSLPIKEYQIIDYFFQPNECSHPLKDDVVKIMPVQKQTRAGQRTRFKAFVAIGDGNGHCGLGVKCA
KEVATAIRGAIISAKLSLIPVRRGYWGNKIGDPHTVPMKVSGKCGSVRIRLVPAPRGTQIVGAPTTK

KMLNFAGIKDCFSSSCGKTKTKGNFLRAIFNALSKTYGYLTPDLWKVTNFDKSPYEEWSDFLETYQ
NLKGIKGTV

>gi|124802649|ref|XP_001347548.1| 40S ribosomal protein S2B, putative [Plasmodium falciparum 3D7]

MSNKKGQSPKEESIAKMLICKVHIGTKNLENKMKRYVYTRAKDGVHIINLAKTYEKLQLAARIIVA
ISNPADVVVVSARPFGSRAVLKFAQYTGAQAIAGRWTPGMLTNQIIQKFTEPRLLIVTDPRTDAQSV
KESAYANIPVIALCDSDSPLEHVDIAIPCNNKGKESIALMYWLLAQEVLYLKGVIPRSEPWNVMVD
MFLWRDPEQFELKNLANEENTPTAPHLIENQYAAEAPYDEWTKKEEWNDNTNEDWKNPIAAEEW

>gi|124810210|ref|XP_001348801.1| 40S ribosomal protein S3, putative [Plasmodium falciparum 3D7]

MSAPISKKRKFINDGVFQAELNEFLARILAEDGYSGVEVRVTPIRTEVIIRATRTREVLGDKGRRIRE
LTSLVQKRFFNKSTNSVELFAERVEHRGLCAMAQAESLRYKLLKGLAVRRACYGVLRHIMESGAK
GCEVIVSGKLRAQRAKSMKFRDGYLISTGEPSKRFVNTATRSAQLKQGVLGIKVKIMLPTAIDTRTG
LTSILPDNISVLEPKTDTVDL

>gi|124505117|ref|XP_001351300.1| 40S ribosomal protein S3A, putative [Plasmodium falciparum 3D7]

MAVGKNKRTSKGKKGGKKKVTDVFTKKEWYDLKAPKMFLVRNFGKTLVTKTIGKKLATDSLKG
RIYEVNLADLNNDEDQAHKKIKLSCDHIINRDCYTDFCGLSITRDKLCSLIRKGYTLIEGHTDVKTL
DNYHLRMFCIAFTKKRQNQTKSTCYAQTSQIKKIRKKMVDIMTAEASKVLLKDLVKKFIPESIGKEI
EKQCKKIYPLQNVLIRKVKILKRPKLDISKLMELHTDPKEESGKNVNALPESKEATNILTAELKH

>gi|258597201|ref|XP_001347741.2| 40S ribosomal protein S4, putative [Plasmodium falciparum 3D7]

MGKGIKKHLKRVNAPSHWMLNKMGGQYAPKTSSGPHKLLESIPLVILLRNRLKYALTFDEVKMILI
QKIVKVDNKVRTDCTFPVGLMDVIHITKSNEYFRLLYDIKGRFVPHRITNEESKYKLCKVKKILLRK
GRLSIAVTHDGRSIPYIHPDVKVNDTVRLDLETGKVLEHLKFQVGSLVMVTAGHSVGRVGVISSIDK
NMGTYDIIHVKDSRNKVFATRLSNVFVIGDNTKPYISLPREKGIKLDIIEERRNRLKALNN

>gi|124511970|ref|XP_001349118.1| 40S ribosomal protein S5, putative [Plasmodium falciparum 3D7]

METTTADIKLFKKWSYEEINIADLSLVDCIAVSQKACVYTPHTAGRYQKKRFRKALCPIVERLVNSM
MMHGRNNGKKLKAIRIVAYAFEIIHLMTGENPLQVFVNAVQKGGPREDSTRIGSAGVVRRQAVDV
SPLRRVNQAIYLICTGARNAAFRNIKSISECLAEEIINCANESSSSYAIKKKDEIERVAKANR

>gi|124513570|ref|XP_001350141.1| 40S ribosomal protein S6, putative [Plasmodium falciparum 3D7]

MKLNISNPLNNVQKSIEIDDEKKLLPFMEKRIGNAVPGDSIGEEFTGYVFRITGGNDKQGFPMIQGV
LTNNRVRLLFKKGMKCYRPRKKGERKRKSVRGCIVGQDLSALNLTLVKKGVNEIPGLTDKAVGKK
LGPKRASKIRKLFNLDKSDDVRKYVIGRAITKNGKTKFIKPKIQRLVTEKRLLRKRNLLQAKEKRR
LEKKQQLKEYKQLLNKYRSELNQQHDVETTKKKKVKKSLSKTNKTASKSKLNTKQEQKDKTEKK
QNKTNNIKNDKSEKKEQAKKKTKTNENTQQTKQNKPDKKNKAKK

>gi|124512798|ref|XP_001349755.1| 40S ribosomal protein S7, putative [Plasmodium falciparum 3D7]

MDAVQKRVLKSNPSDLEKEIAQCLLDIELSSSSDIKTDAKEIKLLSCDLIEVEKLKKKTILIYIPYKIY
TTYVRKIQRKLINELEKKTKKYVVLVAKRTILKGKQKNKSFKIIPRSRTLTSVYDSILEDIVSPSEIIGK
RISMKADGKRVFKIMLDSKERQRDNIEEKLISFAAVYKKITRRDAVFSLPPTNEK

>gi|124808201|ref|XP_001348256.1| 40S ribosomal protein S8e, putative [Plasmodium falciparum 3D7]

MGISRDGRHKLRLTGGKKKIHKKKRKYELGRPPSNTKLGSRQVHVVRGRGRNYKYRAIKLDSGSF
SWPTFGISKNTRIIDVVYNASNNELVRTKTLVKNCIVVIDSHPFTTWYENTFGTTLGKKKKEKKEED
NKEENKQEVENNEEAAKDETTKTYGVIKKIGKSKNIDPLLLEQFKQGRVLACISSRPGQCGKADGY
IIEGDELLFYKRKMDKKKRN

>gi|124506321|ref|XP_001351758.1| 40S ribosomal protein S9, putative [Plasmodium falciparum 3D7]

MPKSYRNYSKTARNPKRPFEKERLDQELKLIGEYGLKNKREIWRVQYLLAKIRSAARYLLTLDEKS
SKRIFQGEALLRRMVRQGLLGENEEKLDYVLGLTLPKLLERRLQTKVFKLGLAKSVHHARVLIRQ
RHIRVGKQMVDIPSFLVRVDSEKHIDFATTSPFGGARPGRVKRKSLKKQKEKTEAEAE

>gi|124511930|ref|XP_001349098.1| 40S ribosomal protein S10, putative [Plasmodium falciparum 3D7]

MDKQTLPHHKYSYIPKQNKKLIYEYLFKEGVIVVEKDAKIPRHPHLNVPNLHIMMTLKSLKSRNY
VEEKYNWKHQYFILNNEGIEYLREFLHLPPSIFPATLSKKTVNRAPKMDEDISRDVRQPMGRGRAF
DRRPFE

>gi|124505009|ref|XP_001351246.1| 40S ribosomal protein S11, putative [Plasmodium falciparum 3D7]

MATTLDVQHERAYQKQEGASFFNSKKIKKGSKSYIRYWKKVGLGFATPKEAKEGVYVDKKCPFTG
NVSIRGRILKGMVISNKMKRTIIIRRNYLHYVKKYNRFEKRHKNIPCHCSPCFDVKEGDIVTVGQCR
PLSKTVRFNVLHVEKHQIFGSARKQFVLF

>gi|124504807|ref|XP_001351146.1| 40S ribosomal protein S12, putative [Plasmodium falciparum 3D7]

MSDVESADNNVVVEEKAVFDNVTAIQKVIKNAHVHDGLKIGIREVIKSIESQEAKVCFLSDVCSEPA
YKKLITTLCAEKNIPLFMVQNDSKDLGHWAGLFKLDNEGNARKIIGASSVAVVDFGEDSAEKDFLL
SQNQTVTA

>gi|124513900|ref|XP_001350306.1| 40S ribosomal protein S13, putative [Plasmodium falciparum 3D7]

MGRMYGKGKGISSSTLPYKRKQPSWLKQKPSEIEDAIIKLAKKGQTPSQIGATLRDNYGIPQVKSV
TGNKILRILRAQGIATTIPEDLYFLIKKAVSMRKHLEKNKKDKDCKFRLILTESKIHRISRYYKRKKLL
PSNWKYQSSTASALIA

>gi|124506243|ref|XP_001351719.1| 40S ribosomal protein S14, putative [Plasmodium falciparum 3D7]

MASKKVKTPQPETAIVSGPQPKEGELVFGVAHIFASFNDTFIHVTDLSGRETLVRITGGMKVKADRD
ESSPYAAMMAAQDVAARLKELGVTAIHIKLRASGGTKSKTPGPGAQSALRALARSGLKIGRIEDVT
PIPTDSTRKKSGRRGRRL

>gi|124504993|ref|XP_001351238.1| 40S ribosomal protein S15A, putative [Plasmodium falciparum 3D7]

MVRMSVLADCLKTINNAEKRGRRQVLIRPSSKVVIKFLQYMQKKGYIGSFEIVDDHRSGKIVVNLL
GRINKCAVISPRYDVKLDEIEKIITSILPSRLFGHLILTTPYGIMDHEEARRKHTGGKVLGFFF

>gi|296005391|ref|XP_002809019.1| 40S ribosomal protein S15/S19, putative [Plasmodium falciparum 3D7]

MEDANKPKKRTFRTFQYRGVDLDKLLDLSQDELIKLFKARQRRKFQRGISKKAKSLLKKIRKSKK
NCEPGEKPNPVPTHLRNMTIIPEMVGSIVAVHNGKQYTNVEIKPEMIGYYLGEFSITYKHTRHGKPG
IGATHSSRFIPLK

>gi|124512506|ref|XP_001349386.1| 40S ribosomal protein S16, putative [Plasmodium falciparum 3D7]

MTTKVKRVQTFGKKKTAVAVATVTNGKGLIKLNGKNLDLVEPYILKTKVYEPLWLIGSGKLKNLDI
RIRVKGGGQTSQIYAIRQAIGKGIISYYQKYVDESTKKELKDVLLRYDRSLLVGDTRRCEPKKFGGK
GARARYQKSYR

>gi|23496944|gb|AAN36495.1| 40S ribosomal protein S17, putative [Plasmodium falciparum 3D7]

MGRVRTKTIKRAARQIVEKYYAKLTLDFQINKKITEEVAIIPSKRMKNKVAGFVTHLMKRIQKGPVR
GISLKLQEEERERRLDFVPEKSQIDVSVIYVEPDTLRMIKSLGINISNMKVHNPMINTNQQKQNRMN
NQF

>gi|124804238|ref|XP_001347943.1| 40S ribosomal protein S18, putative [Plasmodium falciparum 3D7]

MSLQVIDNNDFQHILRILNTNVDGKEKVIIALTAIKGIGKRMATVICKQANVDPTKRAGELTTEEIDN
IVHIMSTPTQFKIPDWFLNRRKDLKEGKNIHVIANQLDSYLREDLERMKKIRLHRGLRHHWGLRVR
GQHTKTTGRRGRTVGVAKKKGA

>gi|124505571|ref|XP_001351527.1| 40S ribosomal protein S19, putative [Plasmodium falciparum 3D7]

MAEQFTEDIGVVNKRLLEPVPFVKTNNCIKDVDADLFIRSYATHLKLHNKITYPKWCTFVKTGKGR
KLAPLNEDWYFIRASSILRRLYLHPDIGVGFLRRQFSSKQRRGVAPNHTSLASGKILRSILQQLENLG
YVEQNPKKKGRRLTTKGENAINNFARYINKKVYNKE

>gi|124801981|ref|XP_001347323.1| 40S ribosomal protein S20e, putative [Plasmodium falciparum 3D7]

MSKLMKGAIDNEKYRLRRIRIALTSKNLRAIEKVCSDIMKGAKEKNLNVSGPVRLPVKTLRITTRK
SPCGEGTNTWDRFELRIYKRLIDLYSQCEVVTQMTSINIDPVVEVEVIITDS

>gi|124804821|ref|XP_001348121.1| 40S ribosomal protein S21e, putative [Plasmodium falciparum 3D7]

MFNDQKVLVDIYIPRKCSATSRLIPAKEHGAVQINVGMVDANGVYNGKTETFAISGHVRQNGESDA
CLNRLMYEKKLLSFQN

>gi|124504805|ref|XP_001351145.1| 40S ribosomal protein S23, putative [Plasmodium falciparum 3D7]

MGSGKPSGLRAARKLRIRRRTQRWADKSYKKSHLGTRWKSNPFRGSSHAKGIVVEKVAIEAKQPN
SAYRKCVRVQLIKNGKKITAFVPGDGCLNFIDENDEVLVSGFGRSGHSVGDLPGVKFKVVKVARVS
LLALFKEKKEKPRS

>gi|124506309|ref|XP_001351752.1| 40S ribosomal protein S24, putative [Plasmodium falciparum 3D7]

MTDQFTIRVKKYMSNPLLRRKQFALEILHPNKGSVAKKEVKERLAKMYKLNNVNTIVLFGFKTLF
GGGRTKGFGLIYKNVDAVKKFEKKYRLVREGLIDKETKAGRRASKELKNRRKKVRGTEKTKVSG
AKKK

>gi|258597702|ref|XP_001348379.2| 40S ribosomal protein S25, putative [Plasmodium falciparum 3D7]

MPPKERKTKEQIAAAAAAASGRTKKKKWGKGKNKEKLNHAVFIDKSLHSKILECKNMKVITPSAIA
EKYKVNLSVARAVINHLADKKLIAEVCVQSHSQKLYTKVA

>gi|124801397|ref|XP_001349683.1| 40S ribosomal protein S26e, putative [Plasmodium falciparum 3D7]

MPKKRRNGGRSKHNRGHVNPLRCSNCGRCVPKDKAIKRFNIRNIVDTSAQRDIKEASVYSTFQLP
KLYIKQCYCVSCAIHSRFVRVRSREQRRVRKETAKHVNPSQL

>gi|124512908|ref|XP_001349810.1| 40S ribosomal protein S27, putative [Plasmodium falciparum 3D7]

MNVDLLNPDPVEESKKHKLKRLIPTPNSYFMDVKCPGCLQITTLFSHAQNVVLCGSCNIMLCQPTG
GKCKLTEGCSFRKKIE

>gi|124810100|ref|XP_001348759.1| 40S ribosomal protein S28e, putative [Plasmodium falciparum 3D7]

MEKSKLAKVEKVLGRTGSRGGVIQVRAQFMGDSELAGRFLIRNVKGPVREGDILALLETEREARR
LR

>gi|296004806|ref|XP_002808755.1| 40S ribosomal protein S29, putative [Plasmodium falciparum 3D7]

MGCILNVHPKKYGQGSRQCRVCSNKHAIIRKYNINICRQCFRERADIIGFKKYR

>gi|124801435|ref|XP_001349693.1| 40S ribosomal protein S30, putative [Plasmodium falciparum 3D7]

MGKVHGSLARAGKVKNQTPKVPKLDKKKRLTGRAKKRQLYNRRFSDNGGRKKGPNSKA

>gi|124808012|ref|XP_001348200.1| 40S ribosomal protein S31/UBI, putative [Plasmodium falciparum 3D7]

MKILINIPYDESLCLESSNINNIKNVKEQIFELKGIPYELQKLYKNGRHLEDEELLEIDKSDYAYTLNL
NFGLLGGAKKKKKKVYKKPKKEKHKKKKVKLAVLKFYKVGDDGKVFRLKRQCDNCAPGTLMA
SHFDRDYCGRCHLTIMKK

>gi|60729641|pir||JC7987 receptor for activated C kinase, RACK protein - Plasmodium falciparum

MMDNIKEAEISLRGVLEGGHSDWVTSVSTPTDPKLKTIVSASRDKKLIVWNINTDDDSGEIGTARK
SLTGHSQAINDVSISSDGLFALSGSWDRSVRLWDLSLGETIRSFIGHTSDVFSVSFSPDNRQIVSASR
DKTIKLWNTLAQCKYTITDQQHTDWITYVRFSPSPNQAIIVSCGWDKLVKVWNLKNCDLNKNLEG
HTGVLNTVTISPDGSLCASGGKDGVAKLWDVKEGKHLYSLETGSTINSLCFSPCDYWLCAATDRFI
RIWNLESKLIISEIYPVKQSKIGVPWCTSLTWSANGQLLYCGSTDGNIYVYEVKKHSV

>gi|124806752|ref|XP_001350823.1| translation initiation factor SUI1, putative [Plasmodium falciparum 3D7]

MNLAIQNLGINDPFTNENIVDKGNGKSNATNLIHIRNQQRNGRKSVTTVQGLGKTFDLKKMVRAL
KKEFNCNGTIIEDIEHGSIIQLQGDKRNNVKEFLIREGICALEHIRIHGA